

[illegible]

QY	1130	GAAGAGTAACTCAATAAACTAAAG---TTTTGAAAAAGAAATCTAAGATGTT-GCA	1188
Db	48	AlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAla	67
QY	1186	TGTCCACAAAAG---AATCATCTACAAAGCAAGTCCCATGATCATCAGG---TCCCATCAG	1239
Db	68	CysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGlu	87
QY	1240	ATCCAAACAAGGAAGATGAAGAATATTTCTTT-GATTCTCGGGTCTCTTTG-----AAG	1292
Db	88	SerLysGlnGluLysAspGluLutyrSerCysAspSerArg---SerLeuPheGluSer	106
QY	1293	TTCTCGAAGATTCAAAGTG---GTATACCTGAGTCTATATATAAAAGTAATGAGATAAT	1349
Db	107	SerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsn	126
QY	1350	AGAAAG---TAGAAAGCCCTCTAAAGCCATCTGCCTTCAAGCGCCATTGAAAGCAAAA	1406
Db	127	ArgGluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln	146
QY	1407	---CTCTTTCCAAATAAGCCTTTG-----AATGAAGATGACAAACATTAAGACA	1454
Db	147	AsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAla	164
QY	1455	CATCCGGT-TTCCCAACCCG---AATCCAAACAAGGACTATAGAAGAAATCTTCGGATTCT	1510
Db	165	AspProMetPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer	184
QY	1511	AGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGTGTTTACCAAGCTACACATCA	1564
Db	185	GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlu	204
QY	1565	AAAAAGATAGATAAAAA---AATCGAAAA---TAGAAGAGTCCTTAATAAGGTGGTCTC	1618
Db	204	nLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuL	224
QY	1619	TGAAGGTACTCTCGGAAT---GAAATTTCTATTCCACTAAAGCCT---AGAATTGAAGAC	1672
Db	224	eLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAsp	243
QY	1673	ATGCAAACTTTCAAAGC-GAGCCTCCGGGAAGC---CACTGCTTCGAGCCTCCACTGAA	1728
Db	244	MetGlnThrPheLysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGlu	263
QY	1729	ATC---AAAAGTCTGTCCCAATAAGACCTTGGAAATGAAATAAAAAATGAGGAGCAGA	1785
Db	264	MetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAla	283
QY	1786	TGAGAA---CTCCCATCAGATCCCAACAAGACACTATAAGAAA---TTCTGGGATACT	1839
Db	284	AspGluLeuLeuProSerGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspThr	303
QY	1840	GGAGTCTCTG---TAGACTGTTTCACAGAAGGATTG-TGTTTACCAGAG---CTCGGCTC	1892
Db	304	GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisGlu	323
QY	1893	AAAAAGATAGATAAAATAATGAAAATTAGAAGGCT---CCCTGTAAAGTGTCCTTCT	1949
Db	323	nLysGluIleAspLysIleAsnGlyLysLeuGluGlySerProGlyLysAspGlyLeuLe	343
QY	1950	GAGGCTAACT---GGGAATCAAAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACA	2003
Db	343	uLysAlaAsnCysGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspM	363
QY	2004	TGCNAATCTCAAAG---CAGGCTCCCGAGAGCCATCTGCCTTCGAGCTCCATTGAAA-	2059
Db	363	etGlnThrPheLysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluM	383
QY	2060	--TCAAAAGTCTGTCAAATTAAC---CTTGGAAATTAACAATGACAAACATTAAGGAG	2114
Db	383	etGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAla	402
QY	2115	ATGAGAA---CTCCCATCA---AATCCAAACAAGGACTATGAGAAAGTCTTCGGATTCT	2168

```

Db      756 LysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLys 775
QY      3149 ---GGATCTCAATATGCGGCACT---AAAGTTCTGATAGCTGAGAAACAATGCTCAT 3202
Db      776 ArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeu 795
QY      3203 CTAATTTGAGAAACACAGACAAAGAAA---CTAGAGCAGAAATTCGAATA-CACC 3255
Db      795 hrSerLysLeuLysGluLysGlnAspLysGluLeuGluAlaGluLeuGluSerHis 815
QY      3256 ATCTGATCGCTTCGCTGATACAGC-----CATCATCAATTTGTACATCAAG 3303
Db      815 isPro-----ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSer 832
QY      3304 AAAAGTAAAG---ACCTGCTTCCACATGAG---GAGATGCTGTTTGCAGAAGAAATGA 3357
Db      832 rGlySerGlnGluProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMet 852
QY      3358 ATGTTGATGTG---AGAGTACGATATTAACAATGAGTGTCTCCATCACCATTTC---TGA 3411
Db      852 snValAspValSerSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGlu 871
QY      3412 GCTCAAGGAATCCAAACCTAAATATCT---CAATTAGCAGGAGATGCTTAAGAG 3468
Db      872 AlaGlnArgLysSerLysLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArg 891
QY      3469 A---AATPACATGTTTCAGAACATCACAAGAGA---CAACGTGAAACACAGTGTCAAT- 3521
Db      891 gGluAsnThrLeuValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGln 911
QY      3522 --GAAGGACGACACATGATCAAAAC-----GACAAGATATGCAACAAACACCTG 3573
Db      911 etLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluG 931
QY      3574 AACAGCAGAGTCTCTAGTCAGAAATATTTCACATACAGCAAGCAAAATTTGCGCTTCACAG 3633
Db      931 lngGlnSerLeu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGln 950
QY      3634 CAATT---AGTCATCCATCAAGAGCTGCAACAACAAAGCAGATAAACAAT----- 3680
Db      951 GlnGlnLeuValHisAlaHisLys-----LysAlaAspAsnLysSerLysIle 966
QY      3681 -----GATATTCATTCTTG---AGAGAAATGCAACACATCTCTCTAAAGAGAAAG 3731
Db      967 ThrIleAspIleHisPheLeuGluArgLysMetGlnHisHisLeuLeuLysGluLysAsn 986
QY      3732 AGGAGATATT---AATTACATACCATTTTAAACCGTATATT---CAATATGAAAGA 3785
Db      987 Glu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGluLysG 1006
QY      3786 GAAAGAGAAACAGAAACTCATG 3807
Db      1006 uLysAlaGluThrGluValIle 1013

RESULT 11
US-10-124-805-553
; Sequence 553, Application US/10124805
; Publication NO. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 553
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

US-10-124-805-553
Alignment Scores:
Pred. No.: 4,54e-80 Length: 1013
Score: 926.50 Matches: 504
Percent Similarity: 59.39% Conservative: 116
Best Local Similarity: 48.28% Mismatches: 293
Query Match: 13.08% Indels: 144
DB: 14 Gaps: 83

US-09-602-362E-22 (1-4115) x US-10-124-805-553 (1-1013)
QY      1018 CACCTAGGAGAAATATGAGTCCCGAAAGAAACACTGAGAAATACGTG---GGCAGAAAA 1074
Db      8 HisValGlySerMetSerProAlaLysGluThrSerGluLysPheThrTTPAlaAlaLys 27
QY      1075 GGAAGACCTAGGAAGAC-GCATGGGAGAAAAAGAAA---CCCTGTAAGATGAGTGGGT- 1129
Db      28 GlyArgProArgLysIleAlaTrpGluLysGluThrProValLysThrGlyCysVal 47
QY      1130 GCAAGAGTAACATCTAATAAACTAAAG---TTTTCAAAAAGGAAATCTAAGATGTT-GCA 1185
Db      48 AlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAla 67
QY      1186 TGTCACAAAAG---AATCATCTCAAGCAAGTGCATGATCAGAG---TCCCATCAG 1239
Db      68 CysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGlu 87
QY      1240 ATCCAAACAAAGGAGAGATGAAGATATTTT-CATTCTCGGTCTCTTTG-----AAG 1292
Db      88 SerLysGlnGluAspGluTyrSerCysAspSerArg---SerLeuPheGluSer 106
QY      1293 TTCTCGCAGATTCAAGT---GTATACCTGAGTCTATATATAAAAGTATATGATATAAT 1349
Db      107 SerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsn 126
QY      1350 AAGAAAG---TAGAAAGCCCTCTAAAGGCATCTCCCTTCAAGCCGCAATTTGAAAGCAAAA 1406
Db      127 ArgGluValGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln 146
QY      1407 ---CTCTTTCAAATTAAGCCTTTG-----AATGAAGATGAACAAACATTTGAAGCA 1454
Db      147 AsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAla 164
QY      1455 GATCCGTG-TTCCCAACCG---AATCCAAACAAAGAGCTATAAGAAATTTCTGGATTCT 1510
Db      165 AspProMetPheProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer 184
QY      1511 AGAGTCTCTG---GAGACTGTTTACAGAGG---ATTGTGTTTACCAAGGCTACACATCA 1564
Db      185 -GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisG 204
QY      1565 AAAAGAATAGATAAAAA---AATGGAATA---TAGAAGAGTCCCTTAATAAGTGGTCTTC 1618
Db      204 nLysGlu-IleAspLysIleAsnGlyLysLeuGluSerProAsnLysAspGlyLeu 224
QY      1619 TGAAGGCTACCTCGGAAT---GAAATTTCTATTCCACTTAAGCCT---AGAATTGAAGAC 1672
Db      224 euLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeuGluLysAsp 243
QY      1673 ATGCAAACTTTCAAAGC-GAGCCTCCGGAAGC---CATCTGCTTCGAGCCTCCACTGAA 1728
Db      244 MetGlnThrPheLysAlaGluProGlyLysProSerAlaPheGluProAlaThrGlu 263
QY      1729 ATC---AAAAGTCTGTCCCAATAAGCCTTGGAAATGAAATGAAATGAAACATGAGGCAGA 1785
Db      264 MetGlnLysSerValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAla 283
QY      1786 TCAGAA---CTCCCATCAGATCCAAACAAAGAGCTATAAGAAAA---TTCTGGGTACT 1839
Db      284 AspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspThr 303
QY      1840 GGAGTCTCTG---TAGACTGTGTTTCAGAGAGGATTG-TGTTTACCAAGG---CTGCGCTC 1892

```


Db 304 -GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisG1 323
QY 1893 AAAAGAAATAGATAAAATTAAGGAAATTAAGAAGGT---CCCTGTTTAAAGTGTCTTCT 1949
Db 323 nLysGluLeuAspLysIleAsnGlyLysLeuGluGlySerProGlyLysAspGlyLeuLe 343
QY 1950 GAGGCTAACT---GGGAATGAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACA 2003
Db 343 uLysAlaAsnCysGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspM 363
QY 2004 TGCAAACTTCAAAG---CAGCGCTCCGAGAGCCATCTCCCTCGAGCCTCCANTGAAA- 2059
Db 363 etGlnThrPheLysAlaGluProGluLysProSerAlaPheGluProAlaIleGluM 383
QY 2060 --TCAAAAGTCTGCCAATAAAC---CTTGGAAATTAAGATGAACAACATTGAGCGAG 2114
Db 383 etGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAla 402
QY 2115 ATGAGAA---CTCCCATCA---AATCCAAACAAAGGACTATGAGAAAGTTCTTGGGATTC 2168
Db 403 Asp-GluIleLeuProSerGluSerLysGlnLysAspTy-GluGluSerSerTrpAspSe 422
QY 2169 TGGAGTCTCTG---TAGAGTCTTTCAGAGAGGATGGTG---TTTACCCAGGCTACATC 2222
Db 422 r-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisG 442
QY 2223 AAAAAGAAATAGATAAAA---TAATGGAAATAGAGAGTCTCT-GATAATGAGTTTCT 2278
Db 442 lNysGluIleAspLysIleAsnGlyLysLeuGluGlySerProAspAsnAspGlyPheL 462
QY 2279 GAAGGCTCCCT---GAGAATGAAGTTCTATTCCACTAAAGCCCT---TGAATTGATGGCA 2332
Db 462 euLysSerProCysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspM 482
QY 2333 TGCAAACTTTCAGAGCAGGCTCCCG---AGAGCCATCTCTTCGAGCCTGCATTGAAAT 2389
Db 482 etGlnThrPheLysAlaGluProGluLysProSerAlaPheGluProAlaIle-Glu 501
QY 2390 ---GAAAGTCTGTCCAAATAAAC---CTTGGAAATTAAGATGAACAACATTGAGACAG 2443
Db 502 MetGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuAr 520
QY 2444 ATCAGAT---TTCCCTTCAGATCAAAACAAAGAGGATTGAGAAA---TTCTGGGA 2494
Db 520 gAlaAspGlnMetPheProSerGluSerLysGlnLysAsnValGluGluAsnSerTrpAs 540
QY 2495 TTCTGAAGTCT---CCGTGACTGTTTCACAGAGGATGGTGTTACC---CAGGCTACA 2548
Db 540 pSerGlu-SerLeuArgGluThrValSerGlnLysasp-ValCysValProLysAlaThr 559
QY 2549 CACAAAGAAGAAAGATAAAT---AGTGGAAATTAAGAGATTCACTAGCCTATCAAAAT 2605
Db 560 HisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLys 579
QY 2606 ---CTTGGTACAGTTCAAT---CTTGTCAAAGGCAAGGAACTTCAAAAGACACTGTGAA 2659
Db 580 lLysLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGlu 599
QY 2660 CA-CTGACAGAAATGGACAAT---GAAAGAGAGTTTGTGTACTGAAAAGAAC--- 2712
Db 600 GlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeu 619
QY 2713 TGTGAGACAAAGAAATAAATACACAGTAGAGAA---CCAAAGTTAAATGGGAACAAGA 2769
Db 620 SerGluAlaLysGluLeuLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnG 639
QY 2770 G---TCTGCACTAGATTGACTTAACCAA---GAAAAGAGAGAGAAATGCCGATA 2823
Db 639 uLeuCysSerValArgLeuThrLeuAsnGlnGluGluGlyArg-ArgAsnAlaAspI 659
QY 2824 TATTAAGAAAAA---TTAGGAAGAAATTAGAGAATPCGAGACGACATAG 2871
Db 1006 uLysAlaGluThrGluValIle 1013

Db 659 leLeuAsnGluLysIleArgGluGluLeuGlyArgIle-GluGluGlnHisArg----- 676
QY 2872 AAAGAGTTAGAACTGAAACACAACTTCAAAG---GCTCTCAGATACAAG---ATAAGAAT 2925
Db 677 LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLe 696
QY 2926 TGAAGGTCTAGAAAGTAATTTGAATAGGTTTCTCA---ACTCATGAAATGAAG---ATTA 2979
Db 696 uLysSerValGluSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrl 716
QY 2980 TTCTTACATCAAAATTCATGTTGAAAAGGAATGGCA---TGCAAAACTGGAATAGCCA 3036
Db 716 euLeuHisGluAsnCysMetLeuLysGluIleAlaMetLeuLysLeuGluIleAlaI 736
QY 3037 CACTA---AACACCAATCCAGGAAAGGAAATAAATC-TTTGAGGACTTAAGATTTTAA 3092
Db 736 hrLeuLysHis-GlnTyrglnGluLysGluAsnLysTyrrPheGluAspIleLysIleLeu 755
QY 3093 AGA---AAAGATGCTCAACTTAG-ATGACCCCTAAAGCTGAAAAGGAAATCATTACTAAAG 3148
Db 756 LysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLys 775
QY 3149 ---GGATCTCAATATGTGGGCAGCT---AAAGTTCTGTAGCTGAGAAACAATGCTCAT 3202
Db 776 ArgAlaSerGlnTyrrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeu 795
QY 3203 CTAAATTTGAGGAAACACACAAAGAAA---CTAGAGCAGAAATTTGAATA-CACC 3255
Db 795 hrSerLysLeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHis 815
QY 3256 ATCCTGACTGGCTTCGCTGTACAAGC-----CATGATCAATTTGTCACATCAAG 3303
Db 815 isPro-----ArgLeuAlaSerAlaValGlnAspHisaspGlnIleValThrSerA 832
QY 3304 AAAAAGTAGA---ACCTGCTTCCACATTGAG---GAGATGCTGTGTTGCAAGAAAAATGA 3357
Db 832 rGlySerGlnGluProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetA 852
QY 3358 ATGTTGATGTG---AGAGTACGATATTAACTAATGAGTGTCTCATCCACATTC---TCA 3411
Db 852 snValaspValSerSerThrIleTyrr-AsnAsnGluValLeuHisGlnProLeuSerGlu 871
QY 3412 GCTAAAGAAATCCAAACCTTAAATATCT---CAATTAGCAGGAGATGCTTAAGAG 3468
Db 872 AlaGlnArgLysSerLysSerLysIleAsnLeuAsnTyrrAlaGlyasp-AlaLeuAr 891
QY 3469 A---AATACATGTGTTTCAGAACATCACAAGAGA---CAACGTGAAAACAGTGTCAAT- 3521
Db 891 gGluAsnThrLeuValSerGluHis-AlaGlnArgaspGlnArgGluThrGlnCysGlnM 911
QY 3522 --GAAGGAAGCGAACACATGTATCAAAAC-----GACAAGATAATGGAACAAACACCTG 3573
Db 911 etLysGluAlaGluHisMetTyrrGlnAsnGluGlnaspAsnValAsnLysHisThrGluG 931
QY 3574 AACAGCAGCTCTCTAGTCAGAAATATTTCACATACAGCAAAATTTGTGGTTCACAG 3633
Db 931 lNlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGln 950
QY 3634 CAATTT---AGTCATGCATAGAAAAGCTGCAACAAAGCAGATAACAAT----- 3680
Db 951 GlnGlnLeuValHisAlaHisLys-----LysAlaaspAsnLysSerLysIle 966
QY 3681 -----GATATTTCATTTCTTG---AGAGGAAATGCAACACATCTCTTAAAGAGAAAAAG 3731
Db 967 ThrIleAspIleHisPheLeuGluArgLysMetGlnHisLeuLeuLysGluLysAsn 986
QY 3732 AGGAGATATT---AATTACAATACCATTTTAAAAACCCCTATATT---CAATATGAAAGA 3785
Db 987 Glu-GluIlePheAsnTyrrAsnAsnHisLeuLysAsnArgIleTyrrGlnTyrrGluLysG 1006
QY 3786 GAAAGAGAAACAGAACTCATG 3807
Db 1006 uLysAlaGluThrGluValIle 1013

RESULT 12

US-09-604-287A-475
 ; Sequence 475, Application US/09604287A
 ; Patent No. US20020064872A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepler, William T.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 ; FILE REFERENCE: 210121.470C7
 ; CURRENT APPLICATION NUMBER: US/09/604,287A
 ; CURRENT FILING DATE: 2000-06-22
 ; NUMBER OF SEQ ID NOS: 489
 ; SOFTWARE: Fast-Seq for Windows Version 3.0
 ; SEQ ID NO 475
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(1002)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-604-287A-475

Alignment Scores:

Pred. No.: 3,76e-79 Length: 1002
 Score: 917.00 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.94% Indels: 145
 DB: 9 Gaps: 83

US-09-602-362E-22 (1-4115) x US-09-604-287A-475 (1-1002)

QY 1031 ATGAGTCCGGAAGAAACACTGAGAA---ATTAGTGGGCGAGA-AAAGGAAGACCTAGG 1086
 Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTpAlaAlaLysGlyArgProArg 20
 QY 1087 AAGAC-GCATGGGAGAGAAAGAAA---CCTCTAAGATGATGGT-GCAAGAGTAAACA 1141
 Db 21 LysIleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThr 40
 QY 1142 TCTAATAAATAAG---TTTTGAAAAGGAAATCTAAGATGTT-GCATGTCCCAAAAG 1197
 Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
 QY 1198 ---AATCATCTCAAGCAAGTCCATGATCAGAGG---TCCCATCATGATCCAAACAGG 1251
 Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
 QY 1252 GAAGATGAAGAATATTCTTT-GATTCTCGGTCCTTTG-----AAGTCTCGAAGATT 1304
 Db 81 GluAspGluLysThrSerCysAspSerArg---SerLeuPheGluSerSerAlaLysIle 99
 QY 1305 CAAGTG---GTATACCTGAGTCTATATAAAGTAATGAGATAAATAAGAAAG---TAG 1358
 Db 100 GlnValCysIleProGluSerIleIleGlnLysValMetGluIleAsnArgGluValGlu 119
 QY 1359 AAGGCTCTCTAAAGCCATCTGCTTCAAGCGCGCATTTGAAGCAAAA---CTCTTTTCCA 1415
 Db 120 GluProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
 QY 1416 AATAAGCCTTTG-----AATCAAGAAATGAACAAATGAGCAGATCCGTG-TTC 1465
 Db 140 AsnLysAlaPheGluLysGlnGluGln-----ThrLeuArgAlaAspPrometPhe 157
 QY 1466 CCACCG---AATCCAAACAAAGGACTATAGAAGAAATCTTGGGATCTTAGACTCTG-- 1520

Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
 QY 1521 -GAGACTGTTTACAGAAGG---ATTGTGTTTACCAAGGCTACACATCAAAAAGAAATAGAT 1576
 Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
 QY 1577 AAAAA---AATGGAATA---TAGAAGAGTCCCTTAATAAAGTGGTCTTCTCAAGGCTACCT 1630
 Db 197 sPlysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
 QY 1631 CGGAAT---GAAATTTCTATCCACTAAAGCCT---AGAATTGAAGACATGCAAACTTTC 1684
 Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
 QY 1685 AAAGC-GAGCCTCCGGGAAGC---CATCTGCTTCGAGCCTCCACTGAATC---AAAAGT 1737
 Db 237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
 QY 1738 CTGTCCCAATAAGCCTTGGAAATGAAAAATGAAAAACATGAGGCAGATGAGAA---CTC 1794
 Db 257 ValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 276
 QY 1795 CCATCAGAAATCCAAACAAAGGACTATAGAAA---TTCTGGGATATCTGAGTCTCTG-- 1849
 Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSer-GluSerLeuCy 296
 QY 1850 -TAGACTGTTTACAGAAGGATG-TGTTTACCAAGG---CTGCGCTCAAAAAGAAATAGA 1904
 Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleAs 316
 QY 1905 TAAATAATGGAAATTTAGAGGGT---CCTGTATAAGTGGTCTTCTGAGGCTAACT-- 1959
 Db 316 PysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLysAlaAsnCy 336
 QY 1960 -GGAAATGAAAGTCTATTCCAACTAAAGCCTAGA---ATTGATGACATCAAACTTCAA 2015
 Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
 QY 2016 AG---CAGCCTCCGAGAGCCATCTGCTTCGAGCCTCCATTTGAAA---TCAAAAGTCT 2069
 Db 356 ysAlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
 QY 2070 GTCCAAATAAAC---CTTGGAAATTAAGAAATGAACAAACATTGAGGCAGATGAGAA---CT 2123
 Db 376 alProAsn-LysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 395
 QY 2124 CCATCA---AATCCAAACAAAGGACTATGAAAGTCTTGGGATCTTGGAGTCTCTG- 2179
 Db 395 uProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuC 415
 QY 2180 -TAGACTGTTTCCAGAAGGATGGT---TTTACCCAGGCTACACATCAAAAAGAAATAGA 2234
 Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleA 435
 QY 2235 TAAAA---TAATGAAATAAGAGTCTCT-GAATAAGAGTGTCTTCTGAGGCTCCCT- 2289
 Db 435 sPlysIleAsnGlyLysLeuGluGluSerProAsnAspGlyPheLeuLysAlaProC 455
 QY 2290 -GAGAATGAAGTCTTATCCACTAAAGCCT---TGAAATTGATGGCATCAAACTTTCA 2344
 Db 455 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
 QY 2345 AAGCAGGCTCCCG---AGAGCCATCTGCTTCGAGCCTCGATTTGAAT---GAAAAGTCT 2398
 Db 475 ysAlaGluProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
 QY 2399 GTTCCAAATAAAC---CTTGGAAATTAAGAAATGAACAACTTGAAGACAGATTCAGAT----- 2450
 Db 495 ValProAsn-LysAlaLeuGluLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
 QY 2451 -TTCCCTTCAGATCAAAACAAAGAGGTTGAGAAA---TTCTGGGATCTCTCAAGTCT- 2505
 Db 513 tPheProSerGluSerLysGlnLys---ValGluGluAsnSerTrpAspSerGlu-SerL 533

QY 2506 --CCGTGAGTCTTTTCACAGAGGATGGTGTGTACC---CAGGTACACACAAAGAA 2560
 Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
 QY 2561 GGATAAAT---AGTGGAAATAGAGATTCACTAGCTATCAAAAT---CTTGGTACA 2614
 Db 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
 QY 2615 GTTCAT---CTTGTGAAGGAGGAACTTCAAAAGACACTGTGACA-CGTACAGGA 2670
 Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
 QY 2671 AAATGGACAAT---GAAAGAGAGTTTGTGTACTGAAAGAAAC---TGTCAACAAA 2724
 Db 593 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 612
 QY 2725 AGAAATAAATCACACTAGAGAA---CCAAAAGTTAAATGGGAAACAGAG---TCTGCAGT 2778
 Db 613 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerVa 632
 QY 2779 GTAGATTGACTTAAACCAA---GAAAGAGAGAGAGAAATGCCATATATTAAGAAA 2835
 Db 632 lArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGlu 652
 QY 2836 AA-----TTAGGAAGATTAGAGATCGAGAGCAGCATAGAAAGATTAGAA 2883
 Db 652 yIleArgGluGluLeuGlyArgile-GluGluGlnHisArg-----LysGluLeuGlu 669
 QY 2884 GTGAACACAACTTGAAG---GCTCTCAGATACAAG---ATAAGATTGAAGGTGTAGA 2937
 Db 670 ValLysGlnGlnLeu-GluGlnAlaLeuA-gileGlnAspIleGluLeuLysSerValGI 689
 QY 2938 AAGTAATTGAATAGTTTCTCA---ACTCATGAATGAAA---ATTATTCTTACATGAA 2991
 Db 689 uSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGlu 709
 QY 2992 ATTGCATGTTGAAAGGAATTGCCA---TGCAAACTGAAATAGCCACACTA---AACA 3045
 Db 709 snCysMetLeuLysGlyGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis- 728
 QY 3046 CCAATCCAGGAAAGGAAATAATC-TTTGAGACTTAAAGATTTTAAAGA---AAAGAT 3101
 Db 729 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysAsn 748
 QY 3102 GCTGAATCTAG-ATGACCTTAAACTGAAAGGAATCATCTACTAAAG---GGATCTCAA 3157
 Db 749 AlaGluLeuGlnMetThrLeuLysLeuLysGluLysSerLeuThrLysArgAlaSerGln 768
 QY 3158 TATGTGGGCAGCT---AAAGTTCTGATAGCTGAGAAACATGCTCATCTTAATTCAGGA 3214
 Db 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeu 788
 QY 3215 AAACAGACAAAGAAA-----CTAGAGCAGAAATTGAATA-CACCATCTGACTGGC 3267
 Db 788 ysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisPro----- 805
 QY 3268 TTGCTGTACAAG-----CATCATCAATTGTGCATCATCAAGAAAGTAAGA- 3314
 Db 806 --ArgLeuAlaSerAlaValGlnAspHisGlnIleValThrSerArgLysSerGlnG 825
 QY 3315 --ACCTGCTCCCATTTAG---GAGATGCTGTTTGCAGAGAAATGAATTTGATGTG- 3368
 Db 825 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspVal 845
 QY 3369 --AGAGTACATATTAAATAGTGTCTCCATCACCCTTTC---TGAGCTCAAAGAAA 3423
 Db 845 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
 QY 3424 TCCAAAACCTTAAATTTATCT---CAATTAGCAGGAGATGCTCTAAGAGA---AATACATT 3477
 Db 865 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 884

QY 3478 GGTTCAGACATCAAAAGAGA---CAACGTGAAAACAGTGTCAAT---GAAGGAAGCG 3531
 Db 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904
 QY 3532 AACACATGTATCAAAAC-----GACAAGATAATGAAACAAACACCTGAACACAGTC 3585
 Db 904 luHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerL 924
 QY 3586 TCTAGTCAGAAATATTCACTACAGCAAAAATTTGTGGCTTCACCAATTT---AGT 3641
 Db 924 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 943
 QY 3642 CATGCACATAGAAAGCTGACAAACAAAGCAGATAACAAT-----GATATT 3686
 Db 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
 QY 3687 CATTTCTTG---AGAGGAAATGCAACACATCTCTCTAAAGAGAAAAGAGGATATT-- 3741
 Db 960 HisPheLeuGluArgLysMetGlnHisLeuLeuLysGluLysAsnGlu-GluIlePh 979
 QY 3742 -AATTACAATACCATTTAAACACCGTATATT---CAATATGAAAAGAGAAA 3789
 Db 979 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGluLysGluLys 996

RESULT 13
 US-09-551-621-475
 ; Sequence 475, Application US/09551621
 ; Publication No. US20030104366A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuqiu, Jiang
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
 ; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
 ; FILE REFERENCE: 210121.470C5
 ; CURRENT APPLICATION NUMBER: US/09/551,621
 ; CURRENT FILING DATE: 2000-04-17
 ; NUMBER OF SEQ ID NOS: 479
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 475
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(1002)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-551-621-475

Alignment Scores:
 Pred. No.: 3,76e-79 Length: 1002
 Score: 917.00 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.94% Indels: 145
 DB: 10 Gaps: 83

US-09-602-362E-22 (1-4115) x US-09-551-621-475 (1-1002)

QY 1031 ATGAGTCCCGAAAGAAACACTGAGAA---ATTAGTGGGCAGA-AAAGGAAGACCTAGG 1086
 Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 20
 QY 1087 AAGAC-GCATGGGAGAAAAGAAA---CCCTGTAAAGATGGATCGGT-GCAAGAGTAGTACA 1141
 Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
 QY 1142 TCTAATAAATAAAG---TTTGTAAAAGGAATCTAAGATGTT-GCATGTCCCAAAAG 1197
 Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60

632	l	argLeuThrLeu	asnGlnGluGluGluLysArg-Arg	asnAlaAspIleLeu	asnGluL	652
2836	AA	-----TTAGGAAGAA	ATTAGAAGAAATCAGAGACGACATAGAAAGAGTTAGAA	2883		
652	y	llea	argGluGluGluGlyArgIle-GluGluGlnHisArg	-----LysGluLeuGlu	669	
2884	GT	AAAAACACAACTTGAAG	---GCTCTCAGATACAAG---ATAAGAATTGAAGGTGTAGA	2937		
670	V	alysGlnGlnLeu-GluGlnAlaLeu	argIleGlnAspIleGluLeuLysSerValG	689		
2938	AA	GTAAATTTGAATAGGTTCTCA	---ACTCATGAAATGAAA---ATTATTCTTACATGAA	2991		
689	u	serAsnLeu	asnGln-ValSerHisThrHisGlu	asnGlu	asnTyrLeuLeuHisGlu	709
2992	AT	TGCATGTTGAAAAAGGAATTC	---TGCAAACTCGAATAGCCACACTA---AACA	3045		
709	sn	CysMetLeuLysLysGluIleAla	MetLeuLysLeuGluIleAlaThrLeuLysHis	728		
3046	CC	ATCCAGGAAGAAAGAAATAATC	---TTTGACGACTTAAGATTTTAAAGA---AAAGAT	3103		
729	G	lnTyrGlnGlnLysGlu	asnLysTyrPheGlu	asnIleLysGluLysasn	748	
3102	G	CTGAACTTAG-ATGACCCT	TAAACTGAAAGGAATCATCTAAAG---CGATCTCAA	3157		
749	A	lGluLeuGlnMetThrLeuLysLeuLysGluGluSer	LeuThrLysArgAlaSerGln	768		
3158	T	ATGTGGCGAGCT---AAAGTCT	GATGCTGAGAACCAATGCTCATCTTAAATTCGAGA	3214		
769	T	yrSerGly-GlnLeuLysVal	leuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL	788		
3215	AAA	CACAGCAAAAGAAA-----CTAGAGC	AGAAATTTGAATA-CACCATCTCTGACTGCG	3267		
788	y	sGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHis	Pro-----805			
3268	TT	CGCTGTACAAGC-----CATGATCA	ATTGTGACATCAAGAAAAGTAAGA-	3314		
806	--	ArgLeuAlaSerAlaValGlnAspHis	aspGlnIleValThrSerArgLysSerGlnG	825		
3315	--	ACCTGCTTCCACATTGAG---GAGAT	GCTGTTTGCAAAAGAAATCAATTTGATGTG-	3368		
825	l	uProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMet	AsnValaspValS	845		
3369	--	AGAGTACGATATTAAACAATGAGTGC	TCCATCACCATTTC---TCAGCTCAACAGAAA	3423		
845	e	rSerThrIleTyr-AsnAsnGluValLeuHisGlnPro	LeuSerGluAlaGlnArgLys	864		
3424	TC	CAAAACCTAAAAATTATCT---CAATT	AGAGAGATGCTCTTAAGAGA---AATACATT	3477		
865	S	erLysSerLeuLysIleAsnLeu	asnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe	884		
3478	G	TTTTCAGAACATCACAAAGAGA---CAAC	GTGAAACACAGTGTCAAT---GAAGGAAGCG	3531		
884	u	ValSerGluHis-AlaGlnArg	aspGlnArgGluThrGlnCysGlnMetCysGluAlaG	904		
3532	AA	CACATGTATCAAAAC-----GACA	AGATAATGGAACAAACACCTCGAACACAGAGTC	3585		
904	l	uHisMetTyrGlnAsnGluGlnAsp	asnValasnLysHisThrGluGlnGlnGluSerL	924		
3586	TC	TACTCGAAATTTTCAACTACAGCA	AAAAATTTGTGGCTTCACAGCAATT---AGT	3641		
924	e	u-AspGlnLysLeuPheGlnLeuGlnSerLys	asnMetTrpLeuGlnGlnGlnLeuVal	943		
3642	C	ATGACATAGAAAGCTCACAAACAA	ACAGATACAAT-----GATATT	3686		
944	H	isAlaHisLys-----LysAla	asnLysSerLysIleThrIleAspIle	959		
3687	C	ATTTCTTG---AGAGGAAATGCAAC	ATCTCTTAAAGAGAAAAGACAGATATT---3741			
960	H	isPheLeuGluArgLysMetGlnHisHisLeuLeuLysGluLys	asnGluLysasnGluLysIlePhe	979		
3742	-	AATTACAATACCATTTAAACCGCTAT	TATT---CAATATGAAAACAGAGAAA	3789		
979	e	asnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnLysGluLys	996			

RESULT 15

```
US-10-076-622-475
; Sequence 475, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 310..429, 522
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-076-622-475
```

Alignment Scores:

```
Pred. No.: 3,76e-79 Length: 1002
Score: 917.00 Matches: 506
Percent Similarity: 59.77% Conservativeness: 112
Best Local Similarity: 48.94% Mismatches: 285
Query Match: 12.94% Indels: 145
DB: 14 Gaps: 83
```

US-09-602-362E-22 (1-4115) x US-10-076-622-475 (1-1002)

```
QY 1031 ATGAGTCCGGAAGAAACACTGAGAA---ATTACGTGGGCAGAAAGGAGACCTTAGG 1086
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 20
QY 1087 AAGAC-GCATCGGAGAGAAAGAAA---CCCTGTAAAGATGATCGT-GCAAGAGTAACA 1141
Db 21 LysileAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY 1142 TCTAATAAACAAG---TTTTGAAAAGGAATCTAAGATGT-GCATCTCCACAAAAG 1197
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetileAlaCysProThrLys 60
QY 1198 ---AATCATCTACAAGCAAGTGCATGATCAGAGG---TCCCATCAGATCCAAAACAGG 1251
Db 61 GluSerSerThrLysAlaSerAlaAsnAspGluArgPheProSerGluSerLysGlnGlu 80
QY 1252 GAAGATGAAGAATATCTTTT-GATTCTCGGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
Db 81 GluAspGluLysSerCysAspSerArg---SerLeuPheGluSerSerAlaLysile 99
QY 1305 CAAGTG---GTATACCTGAGTCTATATATAAAGTAATGAGATAAATGAAG---TAG 1358
Db 100 GlnValCysileProGluSerileTyrGlnLysValMetGluLysAsnArgGluValGlu 119
QY 1359 AAAGCCCTCTTAAAGCCATCTGCCTTCAAGCCGCCATTTGAAGCAAAA---CTCTTTCCA 1415
Db 120 GluProProLysLysProSerAlaPheLysProAlaileGluMetGlnAsnSerValPro 139
QY 1416 AATAAGCCTTTG-----AATGAAGATGAACAAACATTGAACAGATCCGTG-TTC 1465
Db 140 AsnLysAlaPheGluLysLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
QY 1466 CCACCG---AATCCAAACAAGGACATATAAGAAATCTTGGATTTCTAGATCTCTG--- 1520
Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
QY 1521 -GAGACTGTTTACAGAAGG---ATTGTGTTTACCAGGGCTACACATCAAAAAGATAGAT 1576
```

```
Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
QY 1577 AAAAA---AATGAAAA---TAGAAGAGTCCCTAATAAAGTGTCTTCTGAAGCTACCT 1630
Db 197 spLysileAsnGlyLysLeuGluLysProAsnLysAspGlyLeuLysAlaThrC 217
QY 1631 CGGAAT---GAAATTTCTATTCACATAAAGCCT---AGAATTGAAGACATGCAAACTTC 1684
Db 217 ys-GlyMetLysValSerileProThrLysAlaLeuGluLysAspMetGlnThrPhe 236
QY 1685 AAAGC-CAGCTCCGGGAGC---CATCTGCTCGAGCCCTCACTGAAATC---AAAAGT 1737
Db 237 LysAlaGluProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
QY 1738 CTCTCCCAATAAAGCCTTGAATGAAAAATGAAAAACATGGAGGCAGATGAGAA---CTC 1794
Db 257 ValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGluLysLeu 276
QY 1795 CCATCAGAAATCCAAACAAGGACTATAAGAAA---TTCTGGGATCTGAGTCTCTG--- 1849
Db 277 ProSerGluSerLysGlnLysAspTyrGluLysSerTrpAspSer-GluSerLeuCy 296
QY 1850 -TAGACTGTTTACAGAAGGATTG-TGTTTACCAAGG---CTGCGCTCAAAAAGATAGATA 1904
Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluLys 316
QY 1905 TAAATATATGAAATTAAGAGGT---CCCTGTAAAGTGTCTTCTGAGGCTAACT--- 1959
Db 316 pLysileAsnGlyLysLeuGluLysSerProValLysAspGlyLeuLysAlaAsnCy 336
QY 1960 -GGGAATGAAAGTCTTATTCACAACTAAAGCCTAGA---ATTGATGACATGCAAACTTCAA 2015
Db 336 sGlyMetLysValSerileProThrLysAla-LeuGluLysMetAspMetGlnThrPheL 356
QY 2016 AG---CAGCCCTCCGAGAGCCATCTGCTTCGAGCCTCCATTGAAA---TCAAAAGTCT 2069
Db 356 ysAlaGluProGluLysProSerAlaPheGluProAlaileGluMetGlnLysSerV 376
QY 2070 GTCCAAATAAAC---CTTGGAATTAAGATGAACAAACATTCAGGCAGATGAGAA---CT 2123
Db 376 alProAsn-LysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp-GluLysile 395
QY 2124 CCCATCA---AATCCAAAACAAGGACTATGAGAAAGTTCTTGGGATTTCTGGAGTCTCTG- 2179
Db 395 uProSerGluSerLysGlnLysAspTyrGluLysSerTrpAspSer-GluSerLeuC 415
QY 2180 --TAGACTGTTTCCAGAGGATGGTG---TTTACCCAGGCTACACATCAAAAAGATAGA 2234
Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluLys 435
QY 2235 TAAAA---TAATGGAATAAGAGATGCTCT-GATAATGAGGTTTCTGAGGCTCCCT- 2289
Db 435 spLysileAsnGlyLysLeuGluLysSerProAspAsnAspGlyPheLysAlaProC 455
QY 2290 --GAGAATGAAGTTCTATTCACATAAAGCCT---TGAATTCATGGCATGCAAACTTCA 2344
Db 455 ysArgMetLysValSerileProThrLysAlaLeuGluLysMetAspMetGlnThrPheL 475
QY 2345 AAGCAGGCTCCCG---AGAGCCATCTCTTCGAGCCCTGCAATTGAAAT---GAAAAGTCT 2398
Db 475 ysAlaGluProGluLysProSerAlaPheGluProAlaile-GluMetGlnLysSer 494
QY 2399 GTTCCAAATAAAC---CTTGAATTAAGATGAACAAACATTCAGACAGATCAGAT----- 2450
Db 495 ValProAsn-LysAlaLeuGluLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
QY 2451 -TTCCCTTCAGAAATCAAAACAAGAGGTTGAGAAAA---TTCTGGGATTTCTGAAGTCT- 2505
Db 513 tPheProSerGluSerLysGlnLys---ValGluGluAsnSerTrpAspSerGlu-SerL 533
QY 2506 --CGGTGACTGTTTACAGAAGGATGTTGTGTACC---CAGGCTACACACAAAAGAAA 2560
```


Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
 QY 2561 GGATAAAAT---AGTGGAAATTAGAAGATTCTAGCCCTATCAAAAT---CTTGCTACA 2614
 Db 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
 QY 2615 GTTCAT---CTTGAAAGGCGAGGAACCTTCAAAAGACACTGTGAACA-CGTACAGGA 2670
 Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
 QY 2671 AATGGACAAAT---GAAAGAGAGTTTGTGTACTGCAAAAGAAC---TGTACAGACAAA 2724
 Db 593 LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 612
 QY 2725 AGAAATAAATCACAGTAGAGAA---CCAAAAGTTAAATGGGAACAAGAG---TCTGCAGT 2778
 Db 613 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGluLysSerVa 632
 QY 2779 GTGATTGCTTAACCAA---GAAAGAGAGAGAGAAATCCGATATATATAAGAAA 2835
 Db 632 lArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGluL 652
 QY 2836 AA-----TTAGGAAGAAATAGAGAATCGAGAGCAGCATAGAAAGAGTTAGAA 2883
 Db 652 yslleArgGluLeuGlyArgIle-GluGluGlnHisArg-----LysGluLeuGlu 669
 QY 2884 GTGAACACAACTTGAAG---GCTCTCAGATACAAG---ATAGAATTGAAGGTGTAGA 2937
 Db 670 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValG 689
 QY 2938 AAGTAATTTCAATAGTTTCTCA---ACTCATGAATGAAG---ATTATCTTCATCAATGAA 2991
 Db 689 uSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluA 709
 QY 2992 ATTGCATGTTGAAAAAGAAATGCCA---TGCAAAACTGGAAATAGCCACACTA---AACA 3045
 Db 709 snCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis 728
 QY 3046 CCAATCCAGGAAAGGAAATAAATC-TTTGAGGACTTAAGATTTTAAAGA---AAAGAT 3101
 Db 729 GlnTyr-GlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsn 748
 QY 3102 GCTGAACCTTAG-ATGACCCCTAAACTGAAAGGAATCATTAATAAG---GGATCTCAA 3157
 Db 749 AlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGln 768
 QY 3158 TATGTGGGAGCT---AAAGTTCTGATAGCTGAGAAACAATGCTCATCTAAATTTGAGGA 3214
 Db 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThr-MetLeuThrSerLysLeuL 788
 QY 3215 AAAACACACAAGAAAA---CTAGAGCGAGAATTTGAATA-CACCATCTGACTGGC 3267
 Db 788 ysgLulysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisPro-----805
 QY 3268 TTGCTGTACAAGC-----CATGATCAATTTGTGACATCAAGAAAAAGTAAGA- 3314
 Db 806 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 825
 QY 3315 --ACCTGCTTCCCATTTAG---GAGATGCTCTTTGCAAGAAAGAAATGAATCTTCATGTG- 3368
 Db 825 luproAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 845
 QY 3369 --AGAGTACGATATTAAACATGAGTGCTCCATCACCACCTTTC---TGAGCTCAAGGAAA 3423
 Db 845 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
 QY 3424 TCCAAACCTTAAATATCT---CAATTAGCAGAGATGCTTAAGAGA---AATACATT 3477
 Db 865 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 884
 QY 3478 GGTTTCAGAACACACAGAGA---CAACGTGAAACACAGTGTCAAT---GAAGGAAGCG 3531
 Db 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904

QY 3532 AACACATGTATCAAAAC-----GACAAGATAATGGAAACAACACCTGAACAGCAGAGTC 3585
 Db 904 luhisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerL 924
 QY 3586 TCTAGTTCAGAAATTTTCAACTACAGCAAAAAATTTGTGGCTTCACAGCAATT----AGT 3641
 Db 924 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 943
 QY 3642 CATGCACATAGAAAGCTGACACAAACCAACCATTAACAT-----GATATT 3686
 Db 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
 QY 3687 CATTTCTCTTG---AGAGGAAATGCAACACATCTCTTAAAGAGAAAAAGAGAGATATT-- 3741
 Db 960 HisPheLeuGluArgLysMetGlnHisLysLeuLeuLysGluLysAsnGlu-GluIlePh 979
 QY 3742 -AATTACATACCATTTTAAACACCGTATATT---CAATATGAAAAAGAGAAA 3789
 Db 979 easnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGlnLysGluLys 996

Search completed: July 15, 2004, 09:44:59
 Job time : 325.589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:36:32 ; Search time 177.082 Seconds
(without alignments)

13131.604 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086

Sequence: 1 ctactctatcacgaacagc.....gagtggaactccactggaaa 4115

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO spool p/US09602362/runat_15072004_093625_21981/app_query.fasta_1.10325
-DB=A_Geneseq_29Jan04 -OPMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362.cgn.1.1.729 @runat_15072004_093625_21981 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1234	17.4	1341	4 AAB84702 Amino aci
2	1234	17.4	1341	5 ABJ05537 Breast ca
3	1234	17.4	1341	6 ABJ37784 Human tum
4	1234	17.4	1341	6 ABJ37784 Human tum
5	1223	17.3	1349	6 ABJ37788 Human tum
6	926.5	13.1	1013	6 ABJ37783 Human tum
7	917	12.9	1002	4 AAB33351 Human bre
8	917	12.9	1002	5 AAB33351 Human bre
9	917	12.9	1002	6 ABJ37741 Human tum
10	917	12.9	1002	8 ADE44427 Human bre

11	914.5	12.9	1095	4 AAU33357 Human bre
12	914.5	12.9	1095	5 ABG78924 Human bre
13	914.5	12.9	1095	6 ABJ37747 Human tum
14	612.5	8.6	1239	6 ABJ37789 Human tum
15	597	8.4	661	6 ABJ37782 Human tum
16	583.5	8.2	650	4 AAB50263 Human bre
17	583.5	8.2	650	4 AAG65983 Human bre
18	583.5	8.2	650	4 AAU33346 Human bre
19	583.5	8.2	650	5 ABG78913 Human bre
20	583.5	8.2	650	5 ABJ37736 Human tum
21	583.5	8.2	650	8 ADE44421 Human bre
22	582	8.2	743	4 AAU33358 Human bre
23	582	8.2	743	5 ABG78925 Human bre
24	582	8.2	743	6 ABJ37748 Human tum
25	544.5	7.7	1225	6 ABR47547 Breast ca
26	481	6.8	1011	4 AAB84703 Amino aci
27	465	6.6	512	4 AAB84701 Amino aci
28	421.5	5.9	445	4 AAB50249 Human bre
29	421.5	5.9	445	4 AAG65987 Human bre
30	421.5	5.9	445	4 AAU33350 Human bre
31	421.5	5.9	445	5 ABG78917 Human bre
32	421.5	5.9	445	6 ABJ37740 Human tum
33	421.5	5.9	445	8 ADE44425 Human bre
34	416.5	5.9	432	4 AAB50244 Human bre
35	416.5	5.9	432	4 AAG65982 Human bre
36	416.5	5.9	432	4 AAU33345 Human bre
37	416.5	5.9	432	5 AAO15779 Human bre
38	416.5	5.9	432	5 ABG78912 Human bre
39	416.5	5.9	432	6 ABJ37735 Human tum
40	416.5	5.9	432	8 ADE44133 Human bre
41	408.5	5.8	466	4 AAB50248 Human bre
42	408.5	5.8	466	4 AAG65986 Human bre
43	408.5	5.8	466	4 AAU33349 Human bre
44	408.5	5.8	466	5 ABG78916 Human bre
45	408.5	5.8	466	6 ABJ37739 Human tum

ALIGNMENTS

RESULT 1

AAB84702

ID AAB84702 standard; protein; 1341 AA.

XX AAB84702;

XX DT 17-SEP-2001 (first entry)

XX DE Amino acid sequence of a human cancer associated antigen.

XX KW Cancer associated antigen; INGI1; tumour suppressor; cancer; vaccine.

XX OS Homo sapiens.

XX PN WO200147959-A2.

XX PD 05-JUL-2001.

XX PF 29-NOV-2000; 2000WO-US042334.

XX PR 30-NOV-1999; 99US-00451739.

XX PR 24-OCT-2000; 2000US-00602362.

XX (LUDW-) LUDWIG INST CANCER RES.

XX (SLOK) SLOAN KETTERING INST CANCER RES.

XX (CORR) CORNELL RES FOUND INC.

XX Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;

XX WPI; 2001-441706/47.

XX Isolated cancer associated nucleic acid molecule identified by SPREX
PT (serological identification of antigens by recombinant expression
PT cloning) technique, useful in nucleic acid based therapies to treat

587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- 606
1766 ATGAAACATCGAGCAGATGAGAA---CTCCATCAGATCCCAACAAAGGACTATAA 1822
607 ----GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyrG 635
1823 GAAAA---TTCTGGGACTGAGTCTCTG---TAGACTGTTTACAGAGGATG-TGT 1875
625 luGluAsnSerTrpAspThr-GluSerLeuGluValSerGlnLysAspValCys 644
1876 TTACCAAGG---CTGCGCTCAAAAGAAATAGATAAAATAATGGAAGGTTTGAAGGGT--- 1929
645 LeuProLysAlaAlaHisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluGlySer 664
1930 CCTGTGTTAAAGTGTCTTCTGAGGCTAACT---GGGAATGAAAGTTCTATTCCAACTAAA 1986
665 ProValLysAspGlyLeuLeuLysAlaAsnGlyMetLysValSerIleProThrLys 684
1987 GCTTAGA---ATTGATGATCATCAAACTTCAAG---CAGSCCTCCCGAGAGCCATCTGC 2040
685 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 704
2041 CTTGAGGCTCATTGAAA---TCAAAAGTCTGTCCAAATAAAC---CTTGGAAATTAAGA 2094
704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
2095 ATGACAAACATTGAGGAGATGAGAA---CTCCCATCA---AATCCAAACAAAGGACTA 2148
724 snGluGlnThrLeuArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyr 743
2149 TGAGAAAGTCTTGGGATCTGGAGTCTCTG---TAGACTGTTTCCAGAGGATGGTG-- 2203
744 GluGluSerSerTrpAspSer-GluSerLeuGluThrValSerGlnLysAspValCys 763
2204 -TTTACCAGGCTACACATCAAAAGAAATAGATAAAA---TAATGGAATAAGAGATC 2259
763 sLeuProLysAlaThrHisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluGly 783
2260 TCT-GATAATGAGGTTTCTGAAGGCTCCCT---GAGAAATGAAAGTTCTATTCCACTAAA 2315
783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLys 803
2316 GCCT---TCAATGTGATGCTGCAAACTTCAAGGAGGCTCCCG---AGAGCCATCTG 2369
803 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 823
2370 CTTGAGGCTCATTGAAAT---GAAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAG 2423
823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
2424 AATGAACACATTGAGACAGATCAGAT---TTCCCTTCAGAAATCAAAACAAAGAGG 2477
843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
2478 TTGAGAAA---TTCTGGGATTTGAAGTCT---CCGTGGAGTCTTTTCAGAGGATGG 2531
862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
2532 TGTGTACC---CAGGCTACACAAAAGAAAGGATAAAT---AGTGGAAATTAAG 2585
881 aCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
2586 ATTCACTAGCCTATCAAAAT---CTTGGTACAGTTTCAT---CTTGTGAAAAGGCAAGGAA 2639
901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaGluL 921
2640 CTTCAAAAGACACTGTCAACA-CGTACAGAAAATGAACAAAT---GAAAAGAGTTT 2695
921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheC 941
2696 GTGTACTGAAAAGAAC---TGTCAAGACAAAAGAAATAAATCACAGTACAGAA---CCA 2749
941 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln 960

2750 AAAGTTAAATGGCAACAAGAG---TCTGCTAGTGTAGATTGACTTAAACCAA---GAAAAG 2803
961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
2804 AGAAGAGAGAAATGCCGATATATTAAAGAAAAA---TTAGGAAGAATTAG 2851
981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1000
2852 AAGAATCGAGAGCAGCATAGAAAGAGTTAGAAAGTGAACACACAACTTCAAG---GCTCTC 2908
1000 luGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
2909 AGATACAAG---ATAGAATTAAGAGTGTAGAAAGTAATTTGAATAGGTTTCTCA---AC 2962
1018 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
2963 TCATCAAAATGAAA---ATTATTCTTACATGAAATGTCATGTTGAAAAGGAATTGCCA-- 3017
1037 rHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
3018 -TGCAAACTGGAATAGCCACACTA---AACACCAATCCAGGAAAAAGGAAAAATAATC-T 3072
1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTrp 1077
3073 TTGAGGACTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeu 1097
3129 AAAGAAATCATTAATAAG---GGATCTCAATATGTCGGCAGCT---AAAGTTCTGATA 3182
1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
3183 GCTGAGAAACAATGCTCTATTCTAAATTGAGGAAAAACAGACAAAGAAAA---CTAGA 3236
1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluLeuGlu 1136
3237 GCGAATAATTAATA-CACCATCTGACTGCTGCTGCTGCTGTAAGC-----CA 3283
1136 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHis 1153
3284 TGATCAATGTGACATCAAGAAAAAGTAAGA---ACCTGCTTCCACATTCAG---GAGAT 3337
1153 aspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
3338 GCTGTTTCCAAAGAAATGAATGTTGATGTG---AGAGTACGATATTAAACAATGAGTCT 3394
1173 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluValL 1193
3395 CCATCACCACCTTC---TGAGCTCAAAAGAAATCCAAAACCTAAAATTTATCT---CAAT 3448
1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
3449 TAGCAGGAGATGTCTAAGAGA---AATACATTTGGTTTCAGAACATCAACAAGAGA---CA 3502
1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGlu 1232
3503 ACCTGAAAACAGTGTCAAT---GAAGGAGCGACACATCTATCAAAAC-----GACAA 3553
1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
3554 GATAATGGAACAAAACCTGAAACAGCAGCTCTTAGTCAGAAATTTATTTCAACTACAAG 3613
1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
3614 CAATAATTTGGCTTCACGCAATT---AGTCATGCACATAGAAAGCTGACAAACAAG 3669
1272 erLysAsnMetTrpLeuGlnGlnLeuValHisAlaHisLys-----LysA 1288
3670 CAGATAACAAT---GATATTCAATTTCTTTG---AGAGGAAATGCAACAC 3711
1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHis 1308

QY 3712 ATCTCTAAAGAGAGAGAGAGATATT---AATTACATACCATTTTAAAGACGGTAT 3768
 Db 1308 isleuLeuGluGluysaenGlu-GluilePheAsnTyraAsnHisleuLysAsnArg 1327
 QY 3769 ATT---CAATATGAAAGAGAGAAA 3789
 Db 1328 ileTyGlnTyGluLysGluLys 1335

RESULT 2

ABJ05537
 ID ABJ05537 standard; protein; 1341 AA.
 XX
 AC ABJ05537;
 DT 14-NOV-2002 (first entry)
 XX
 DE Breast cancer-associated protein 2.
 XX
 KW Breast cancer; breast cancer-associated gene sequence; drug development;
 KW pharmacogenetics; biosensor development.
 XX
 OS Unidentified.
 XX
 PN WO200259377-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 24-JAN-2002; 2002WO-US002242.
 XX
 PR 24-JAN-2001; 2001US-0263965P.
 PR 02-FEB-2001; 2001US-0265928P.
 PR 09-APR-2001; 2001US-00829472.
 PR 09-APR-2001; 2001US-0282699P.
 PR 04-MAY-2001; 2001US-0288590P.
 PR 29-MAY-2001; 2001US-0294443P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Mack DH, Gish KC, Afar D;
 XX
 DR WPI; 2002-583738/62.
 DR N-PSDB; ABT07694.
 XX

XX
 PT Detecting a breast cancer-associated transcript in a patient's cell,
 PT useful for diagnosing breast cancer, comprises contacting a biological
 PT sample with a polynucleotide that selectively hybridizes with breast
 PT cancer nucleic acids.
 XX
 PS Disclosure; Page 348-349; 414pp; English.
 XX

XX The invention comprises a method of detecting a breast cancer-associated
 CC transcript in a cell from a patient. The method of the invention involves
 CC contacting a biological sample from the patient with a nucleotide that
 CC hybridizes to one of the 69 breast cancer-associated gene sequences shown
 CC in the specification. The method of the invention is useful in the
 CC diagnosis or prognosis of breast cancer, and for detecting genes that are
 CC up or down-regulated in breast cancer cells. Genes identified by the
 CC method of the invention can be used in diagnostic purposes and also as
 CC targets for screening for therapeutic compounds that modulate breast
 CC cancer (e.g. hormones or antibodies). Identification of genes that are
 CC over or under expressed in breast cancer can additionally provide high
 CC resolution, high-sensitivity datasets which can be used in the areas of
 CC diagnostics, therapeutics, drug development, pharmacogenetics, protein
 CC structure and biosensor development. Amino acid sequences ABJ05536 -
 CC ABJ05604 represent the proteins encoded by the 69 breast cancer-
 CC associated genes of the invention.
 XX

SQ Sequence 1341 AA;

Alignment Scores:

Pred. No.: 1.04e-131 Length: 1341
 Score: 1234.00 Matches: 679
 Percent Similarity: 59.18% Conservative: 146

Best Local Similarity: 48.71% Mismatches: 376
 Query Match: 17.41% Indels: 207
 DB: 5 Gaps: 115

US-09-602-362E-22 (1-4115) x ABJ05537 (1-1341)

QY 96 CAAAGGGAAGAGACATCAACCTTATATACAGAG---CCAGAGAGAGCTGCTCTAAC 152
 Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
 QY 153 TGGGCTGTGTCATGCGCTT-----CAGGAAGTAGTAACATTTCT---GGTA 194
 Db 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuVala 38
 QY 195 GACAGAGTCCAGCTGACGCTCTTG---TGCGGAACACGACACCTCTGTAGAGGC-T 250
 Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
 QY 251 TACAATGCCACAGGAGGCTTGC---AAATATTGATAGATTC-GGTCCGATATAAATC 306
 Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuLeuAspSerGlyAlaAspIleAsnL 78
 QY 307 TCGTG---ATGTGTATGCAACATGGCTTCCATTATGCGTTTATAGTGGATTTGTCAGT- 362
 Db 78 euValAspValTyGlyAsnMetAlaLeuHisTyAlaValTySerGluIleLeuServ 98
 QY 363 --GGTGGCAAACTGCTGTC-CATGTGCAATGCAAGTGC---CAACAAGGCTCCCTCA 416
 Db 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
 QY 417 CACCACTTTTACTACCAT---AACGAAAGAGTGAAGTGTGGAATTTT---GCTGAT 470
 Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle 137
 QY 471 AAAATGCAATGCGAATGCGTTA---ATAAGTTAAATGCGACACCTCATGCTCTGTAT 527
 Db 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
 QY 528 G---TCTGGATCATCAGAGATAGTTGTCATGCTTCTTACGAAA---TGTGACGCTTTT-- 579
 Db 157 iCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
 QY 580 -GTGCAGATATAGTGGACTAAGTCCAGACAT---ATGCTGTTACTGTGATTTCTCACA 635
 Db 177 alalaaspIle-CysGlyValThrAlaGluHisTyAlaValThrCysGlyPheHisIle 197
 QY 636 TTCATAACAAA---TTATGAATATATACGAAATTTACAAATCA---TCAAAATACCA 689
 Db 197 leHisGluGlnIleMet-GlutylIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
 QY 690 T---CAGAGGAAGTCTGCAGAAACCTGATGAGGTGCACCTT---GGCGGAAAGACAC 743
 Db 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
 QY 744 CTGACACCTCAAAAGCT---GGTGGAAAAACACCTGATGAGGTGCACCTT---GGTGG 797
 Db 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
 QY 798 AAGACACTGCAGCGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGTGCATCTCT--- 851
 Db 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
 QY 852 TGGTGGAGGACATCTCACAATTTCAATGTTGGGAAGG---ACATCTCGAAGTTCGA 908
 Db 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 295
 QY 909 ACGTCAGC---AGAAAAACACCTAGGAATTCAGCATCC---TGCAAAAAACATCTGAGA 962
 Db 295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
 QY 963 AATTTCAG---GGCCACAAAGGACCTAGAGATCGCAGGAGAA---AAAGAGAC 1016
 Db 315 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334


```

Db      1037  thisGluAsnGluAsnTyrLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
QY      3018  -TGCAAACTGCAATAGGCACACTA---AACACCAATCCAGGAAAGGAAATTAATC-T 3072
Db      1057  tieuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
QY      3073  TTGAGGACTTAAGATTTTAAAGA---AAGATGCTGAACTTAG-ATGACCCCTAAACTGA 3128
Db      1077  heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1097
QY      3129  AAAGGAATCATTAATAAG---GCATCTCAATATGCGGAGCT---AAGTTCTGATA 3182
Db      1097  ysgLgluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
QY      3183  GCTGAGAAACAATGCTCAATTCATAATTGAGGAAACACAGACAAAGAAA-----CTAGA 3236
Db      1117  AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1136
QY      3237  GGCAGAAATTGAATA-CAACATCTGACTGCTGCTGCTGTAACAAGC-----CA 3283
Db      1136  uAlaGluIleGluSerHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
QY      3284  TCATCAATGTGACATCAAGAAAAGTAAGA---ACCTGCTCCACATTGAG---GAGAT 3337
Db      1153  sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
QY      3338  CCGTGTTCGAAAGAAATGAATGTGATGTG---AGAGTACGATATTAACAATGAGTGCT 3394
Db      1173  aCysLeuGlnArgLysMetAsnValAspValSerThrIleTyr-AsnAsnGluValL 1193
QY      3395  CCATCACCATTTTC---TGAGCTCAAGAAATCCAAACCTAAATATATCT---CAAT 3448
Db      1193  euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
QY      3449  TAGCAGGAGATGTCTTAAGAGA---AATACATTTGTTTCAGAAACATCACAAAGAGA---CA 3502
Db      1213  yzAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
QY      3503  ACGTGAACACAGTGTCAAT---GAAGGACGCAACATGATCAAAAC-----GACAA 3553
Db      1232  nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
QY      3554  GATAATGGAAACAAACCTCAACAGCAGAGCTCTAGTCAGAAATATTTCACATACAAG 3613
Db      1252  nValAsnLysHisThrGluGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY      3614  CAAAATTTGGCTTCACGCAATT---AGTCATGCACATGAAAGCTGCAACAAAG 3669
Db      1272  erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY      3670  CAGATAACAT-----GATATTCAATTTCTTG---AGAGGAATGCAACAC 3711
Db      1288  laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
QY      3712  ATCTCTAAAGAAAGAAAGAGAGATATT---AATTACAATACCATTTTAAAGAACCGTAT 3768
Db      1308  isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
QY      3769  ATT---CAATATCAAAAGAGAAA 3789
Db      1328  IleTyrGlnTyrGluLysGluLys 1335

```

RESULT 3

ABJ37784

ID ABJ37784

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

Human; vaccine; gene therapy; T cell stimulation; T cell expansion; tumour; breast cancer; cancer; immune response stimulation.

Homo sapiens.

WO200283956-A1.

24-OCT-2002.

15-APR-2002; 2002WO-US012378.

13-APR-2001; 2001US-00834759.

07-DEC-2001; 2001US-00007805.

13-FEB-2002; 2002US-00076622.

(CORI-) CORIXA CORP.

Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;

Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;

Vedwick TS, McNeill PD, Durham M;

WPI; 2003-103376/09.

New polypeptide and polynucleotide useful for stimulating and/or expanding T cells specific for a tumor protein and treating breast cancer.

Example 9; Page 342-346; 375pp; English.

The invention comprises a method of stimulating and/or expanding T cells specific for a tumor protein. The invention further comprises human nucleic acids and proteins that are associated with tumours (e.g. breast cancer). The method and sequences of the invention are useful for stimulating and/or expanding T cells specific for a tumor protein, detecting the presence of cancer, stimulating an immune response in a patient and treating breast cancer. The present amino acid sequence represents a human tumour-related protein

Sequence 1341 AA;

Alignment Scores:

Pred. No.:	1.04e-131	Length:	1341
Score:	1234.00	Matches:	679
Percent Similarity:	59.18%	Conservative:	146
Best Local Similarity:	48.71%	Mismatches:	376
Query Match:	17.41%	Indels:	207
DB:	6	Gaps:	115

US-09-602-362E-22 (1-4115) x ABJ37784 (1-1341)

QY 96 CAAGGGGAGAGACATCAACCTTAATATACAAGAG---CCGAGAGAGACTGCTCTAAC 152

Db 3 LysargLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21

QY 153 TGGGCTCGTCAATGGGCT-----GAGGAAGTAGTAACATTTC---GGTA 194

Db 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 38

QY 195 GACAGAAGTCCAGCTACGCTCCTTG---TGGCGAACACGAGACACCTCTGATGAAGC-T 250

Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58

QY 251 TACAATGCCACAGGAGGCTTTGC---AAATATTGATAGATTTC-GGTGCCGATATAAATC 306

Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuleaspSerGlyAlaAspIleAsnL 78

QY 307 TCGTG---ATGTGATGCAACATGGCTTCATTTCGTTTATAGTGGATTTTGTGAGT- 362

Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluLeuLeuSerV 98

QY 363 --GGTGCAAACTGCTGTC-CATGGTGCAAGTATCGAAGTGC---CAACAAGGCTGCTCA 416

Db 98 alValalalysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 118

Db 823 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
 QY 2424 AATGACACATTTGAGACAGATCAGAT-----TTCCCTTCAGAAATCAAAACAAAGAGG 2477
 Db 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
 QY 2478 TTGAGAAA---TTCTGGGATCTCGAGCTCT---CCGTGGAGCTCTTCACAGAGGATGG 2531
 Db 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
 QY 2532 TGTGTACC---CAGGCTACACACAAAAGAGGATAAAT---AGTGGAAAATTAGAAG 2585
 Db 881 alCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
 QY 2586 ATTCTAGCTATCAAAAT---CTTGTACAGTTTCAT---CTTGTGNAAGGCAAGGAA 2639
 Db 901 pSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaGluL 921
 QY 2640 CTTCAAAAAGACACTGTGAACA-CGTACAGAAAATGGAACAAT---GAAAAGAGTTT 2695
 Db 921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 941
 QY 2696 GGTGACTGAAAAGAAC---TGTCAGAACAAAAGAAATAATCAGTAGAGAA---CCA 2749
 Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln 960
 QY 2750 AAAGTTAAATGGAAACAAGAG---TCTGCAGTGTAGATTGACTTAAACCAA---GAAAAG 2803
 Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgIleThrLeuAsnGlnGluGlu 980
 QY 2804 AGAAGAGAAATGCCGATATATTAAAGAAAAA-----TTAGGAAGAAATTAG 2851
 Db 981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1000
 QY 2852 RAGATCAGAGCAGCATAGAGAGGTAGAGTGAAGTGAACACAACTGAAG---GCTCTC 2908
 Db 1000 luGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
 QY 2909 AGATACAAAG---ATAAGAAATTGAAGGTGAGAAAGTAAATTTGAATAGTTTCTCA---AC 2962
 Db 1018 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
 QY 2963 TCATGAATGAAA---ATTATCTTACATGAATTCATGTTGAAAAGGAAATGCGCA-- 3017
 Db 1037 rHisGluAsnGluAsnTrpLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
 QY 3018 -TGCAAACTGGAATAGCCACACTA---AACACCAATCCAGAAAAGGAAAATAATC-T 3072
 Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTrpGlnLysGluAsnLysTrp 1077
 QY 3073 TTGAGGACTTAAGATTTTAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
 Db 1077 heGluAspIleLysIleLysLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeu 1097
 QY 3129 AAAGGAATCATTTACTATAAG---GGATCTCAATATGTGGCAGCT---AAAGTTCTGATA 3182
 Db 1097 ysGluLysSerLeuThrLysArgAlaSerGlnTrpSerGly-GlnLeuLysValLeuLe 1116
 QY 3183 GCTGAGAACAAATGCTCATTTCTAATTGAGAAAACAGACAAAGAAA-----CTAGA 3236
 Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1136
 QY 3237 GGCAGAAATTTGAATA-CACCATCTCTGACTGCTTCGCTGACAAAGC-----CA 3283
 Db 1136 uAlaGluLeuGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHI 1153
 QY 3284 TGATCAATTTGACATCAAGAAAAGTAAGA---ACCTGCTTCCACATTGAG---GAGAT 3337
 Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
 QY 3338 GCTGTTTCAAGAAAATGATGTTGATGTC---AGAGTACGATATTAAACATGAGTCT 3394

Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluValL 1193
 QY 3395 CMTACCACTTTC---TGAGCTCAAGAAATCCAAACCTTAAATATATCT---CAAT 3448
 Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysLeuLysIleAsnLeuAsnT 1213
 QY 3449 TAGCAGCAGATGCTTAAGAGA---AATCATTTGGTTTCAGAAACATCACAAGAGAGA---CA 3502
 Db 1213 yAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
 QY 3503 ACCTGAAAACAGTGTCAAT---GAAGGAAGCAACACATGTATCAAAAC-----GACA 3553
 Db 1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTrpGlnAsnGluGlnAspAs 1252
 QY 3554 GATAATGGAACAAACACACCTGAACAGCAGAGTCTCTAGTCAGAAATTTATTTCACTACAAG 3613
 Db 1252 nValAsnLysHisThrGluGlnGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlnS 1272
 QY 3614 CAAAAATTCGTGCTTCACAGCAATT---AGTCATGCATAGAAAAGCTGCACAAAG 3669
 Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
 QY 3670 CAGATAACCAAT-----GATATTCATTTCTTG---AGAGGAAAATGCAACAC 3711
 Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
 QY 3712 ATCTCCTAAAAGAGAAAAGAGAGATATT---AATTACAATPACCATTTAAACAAACCGTAT 3768
 Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTrpAsnAsnHisLeuLysAsnArg 1327
 QY 3769 ATT---CAATATCAAAAAGAGAAA 3789
 Db 1328 IleTyrGlnTrpGluLysGluLys 1335
 RESULT 4
 ID ABR47548 standard; protein; 1341 AA.
 XX ABR47548;
 DT 12-JUN-2003 (first entry)
 DE Breast cancer associated protein sequence SEQ ID NO:334.
 XX Human; breast cancer; cytostatic; gene therapy.
 XX Homo sapiens.
 OS WO2003004989-A2.
 XX 16-JAN-2003.
 XX 21-JUN-2002; 2002WO-US019669.
 XX 21-JUN-2001; 2001US-0299887P.
 PR 27-JUN-2001; 2001US-0301572P.
 PR 18-JUL-2001; 2001US-0306501P.
 PR 25-SEP-2001; 2001US-0325002P.
 PR 05-MAR-2002; 2002US-0362585P.
 PR 14-MAY-2002; 2002US-0380391P.
 XX (MILL-) MILLENIUM PHARM INC.
 PA Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamathar S;
 PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
 PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
 XX WPI; 2003-210381/20.
 DR N-PSDB; ACC50246.
 XX Breast cancer diagnosis or treatment by comparing the level of expression
 PT of a marker in a patient sample with that in the control non-breast
 PT cancer sample.

XX Claim 1; SEQ ID NO 334; 128pp; English.

CC The present invention describes a method for assessing whether a patient
 CC is afflicted with breast cancer. The method comprises comparing the level
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
 CC ABR47386 to ABR47632) in a patient sample and the normal level of
 CC expression of the marker in a control non-breast cancer sample, where a
 CC significant increase in the level of expression of the marker in the
 CC patient sample and the normal level is an indication that the patient is
 CC afflicted with breast cancer. The breast cancer associated sequences from
 CC the present invention have cytotatic activities and can be used in gene
 CC therapy. The method is useful for diagnosing and treating breast cancer.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1341 AA;

Alignment Scores:
 Pred. No.: 1.04e-131 Length: 1341
 Score: 1234.00 Matches: 679
 Percent Similarity: 59.18% Conservative: 146
 Best Local Similarity: 48.71% Mismatches: 376
 Query Match: 17.41% Indels: 207
 DB: 6 Gaps: 115

US-09-602-362E-22 (1-4115) x ABR47548 (1-1341)

QY 96 CAAGGGAAGACACATCACTTATATACAGAG---CCCAGAGAGACTGCTCTAAC 152
 Db 3 LysArgLysThrIleAsnLeuLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
 QY 153 TGGGCGCTGTGCAATGGCCT-----GAGGAGTAGTAAACATTC---GGTA 194
 Db 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuVala 38
 QY 195 GACAGAGTCCGAGTCACTGCTCTTGG---TGGCGACACGACCTCTGATGAAGC-T 250
 Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAla 58
 QY 251 TACAATCCACAGAGGCTTTGTC---AAATATTGATAGATTC-GGTGCCGATATAAATC 306
 Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuLeuAspSerGlyAlaAspIleAsn 78
 QY 307 TCGTG---ATGTGATCAACATGGCTTCATTATGCTTTATAGTGATTTGTGAGT- 362
 Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
 QY 363 --GGTGGCAACTGCTGTC-CATGGTCAGTATCGAAGTGC---CAACAGGCTGCTCA 416
 Db 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
 QY 417 CACCACCTTTTACTACCAT---AACGAAAGAGTGAGCAATTGTGGAATTT---GCTGAT 470
 Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
 QY 471 AAAATGCAATGCAATGGATGGTTA---ATAAGTTAATGACACCCCTGCTGCTGAT 527
 Db 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAla 157
 QY 528 G---TCGTGATCATCAGAGATGATGCTGCTCTTAGCAAAA---TGTGACGCTCTTT-- 579
 Db 157 iCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
 QY 580 -GTGAGATATAGTGGAGTAACCTCAGAACAT---ATGCTGTACTGTGGATTTCTCACA 635
 Db 177 aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisI 197
 QY 636 TTCATACAAA---TTATGATATATACGAAATTTACAAATCA---TCAATATACAA 689
 Db 197 leHisGluGlnIleMet-GlutryIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216

QY 690 T---CAGAAAGGAACCTCTGCGAGAAACCTGATGAGCTGCACCTC---GGCGGAAAGACAC 743
 Db 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
 QY 744 CTGACACCTCAAGACT---GGTGAAGAAACACCTGATGAGCTGCACCTC---GGTGA 797
 Db 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
 QY 798 AAGACACCTGACAGCTGAAAGCT---GGTGAAGAAACACCTGATGAGCTGCATCT--- 851
 Db 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
 QY 852 TGGTGGAGAACATCTGACAAATTCATGTTGGGAAGG---ACATCTGGAAGTTGGA 908
 Db 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheG 295
 QY 909 ACCTGACG---AGAAAAACACCTAGGAATACCATCC---TGCAAAAAACATCTGAGA 962
 Db 295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGlu 315
 QY 963 AATTTACG---GGCCAGCAAGAGACCTAGAGATCGCAGGAGAA---AAAGAGAC 1016
 Db 315 ysPheThrTrpPro-AlaLysGlyArgProArgIlysIleAlaTrpGluLysLysGluAsp 334
 QY 1017 ACACCTAGGGAATATGATCCGAAAGAAACACTGAGAAA-----TTA 1061
 Db 335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
 QY 1062 COTGGCGCAAAAGGAAGACCTAGGAGACCGCATGGGAGAAAGAA---CCCTGTAA 1118
 Db 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
 QY 1119 GATGATCGGT-GCAGAGTAACTCTAATAAATACTAAG---TTTGAAGAAAGGAATCT 1174
 Db 372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
 QY 1175 AAGATGTT-GCATGTCCCAAAAG---AATCATCTACAAAGCAAGTGCATGATCAGAGG 1230
 Db 392 LysMetIleAlaCysProThrLysGluSerThrLysAlaSerAlaAsnAspGlnArg 411
 QY 1231 ---TCCCATCAGATCCAAACAGGAGGAGATGAAGATATTTCTT-GATTCTCGGTCTCT 1286
 Db 412 PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser 430
 QY 1287 TTG-----AGTTCTGCAAGATTCAGTG---GTATACCTGAGTCTATATAAAGT 1337
 Db 431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
 QY 1338 AATGAGATAAATAAGAG---TAGAAAGCCTCTAAAGCCATCTGCTTCAAGCCGCA 1394
 Db 451 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 470
 QY 1395 TTGAAGCAAAA---CTCTTTCCAATAAGCCTTTG-----AATGAAGATGAACA 1442
 Db 471 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
 QY 1443 AACATTGAAGCAGATCCGCTG-TTCCACCG---AATCCAAACAAAGACTATAAGAAAT 1498
 Db 489 ThrLeuArgAlaAspPrometPheProProGluSerLysGlnLysAspTyrGluGluAsn 508
 QY 1499 TCTTGGGANTCTAGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGCTTTACCAA 1552
 Db 509 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProly 528
 QY 1553 GCTACACATCAAAAGAAATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT 1606
 Db 528 salThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL 548
 QY 1607 AAGATGCTCTCTGAGGCTACTCGAAT---GAAATTTCTATTCCCTAAAGCT--- 1660
 Db 548 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 567
 QY 1661 AGAATTGAAGACATGCAAACTTTCAAAGCGAGCCTCC---GGGAAGCCATCTGCT---TC 1714

Db 568 LeuLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG1 587
:::|||||
Qy 1715 GAGCTCCACCTGAATCAAAAGTCTGCTCCCAATAAGCTTGG-----AATGAAAA 1765
|||
Db 587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- 606
|||
Qy 1766 ATGAAAAACATGGAGGACAGATGAGAA---CTCCCATCAGAAATCCAAACAAAGGACTATAA 1822
|||
Db 607 ---GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyrG 625
|||
Qy 1823 GAAAA---TTCTGGGATCTAGGCTCTG---TAGACTGTTTCACAGAGGATG-TGT 1875
|||
Db 625 LuGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 644
|||
Qy 1876 TTACCAAGG---CTCGCTCAAAAGAAATAGATAAATATGGAATAGAGGT--- 1929
|||
Db 645 LeuProLysAlaAlaHisGlnLysGluLeuAspLysAlaAsnGlnLysLeuGluGlySer 664
|||
Qy 1930 CCCTGTTAAAGTGTCTTCTGAGGCTAACT---GGGAATGAAAGTCTTATTCCAACTATAA 1986
|||
Db 665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerLysProThrLys 684
|||
Qy 1987 GCCTAGA---ATTGATGACATGCAAACTTCAAAG---CAGGCTCCCGAGAGCAATCGC 2040
|||
Db 695 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 704
|||
Qy 2041 CTTCGAGCTCCATTGAAA---TCAAAAGTCTGTCCAAATAAAC---CTTGGAAATTAAGA 2094
|||
Db 704 aPheGluProAlaLeuGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
|||
Qy 2095 ATGAAACAACTTGAAGCAGATGAGAA---CTCCCATCA---AATCCAAACAAAGGACTA 2148
|||
Db 724 snGluGlnThrLeuArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyr 743
|||
Qy 2149 TGAGAAAGTCTTCGGGATCTGAGTCTCTG---TAGACTGTTTCAGAGGAGTGTG-- 2203
|||
Db 744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 763
|||
Qy 2204 -TTTACCAGCTACATCAAAAGAAATAGATAAAA---TAATGGAATAAGAGAGTC 2259
|||
Db 763 sLeuProLysAlaThrHisGlnLysGluLeuAspLysAlaAsnGlnLysLeuGluGlu 783
|||
Qy 2260 TCT-GATAATCAGGTTTCTGAGGCTCCCT---GAGAATGAAAGTCTTATTCACACTAA 2315
|||
Db 783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerLysProThrLy 803
|||
Qy 2316 GCCT---TGAATTGATGGCATGCAAACTTTCAAAGCAGGCGCTCCCG---AGAGCCATCTG 2369
|||
Db 803 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 823
|||
Qy 2370 CTTCGAGCTGCATTGAAAT---GAAAGTCTGTTCCAAATAAC---CTTGGAAATTAAG 2423
|||
Db 823 aPheGluProAlaThrHisGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
|||
Qy 2424 AATGAACAACTTGAAGACAGATCAGAT-----TTCCCTTCAGAAATCAAAACAAAGAGG 2477
|||
Db 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
|||
Qy 2478 TTGAGAAAA---TTCTGGGATCTGAAGTCT---CGGTGACTGTTTCACAGAGGATGG 2531
|||
Db 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
|||
Qy 2532 TGTGTACC---CAGGCTACACAAAGAAAGAGGATAAAAT---AGTGGAAATTAAGAAG 2585
|||
Db 881 alCysValProLysAlaThrHisGlnLysGluMetAspLysLysSerGlyLysLeuGluA 901
|||
Qy 2586 ATTCACTAGCTTCAAAAT---CTTGGTACAGTTCAAT---CTTGTGAAAGCAAGGAA 2639
|||
Db 901 spSerThrSerLeuSerLysLysLeuAspThrValHisSerCysGluArgAlaGluL 921
|||
Qy 2640 CTTCAAAACACACTGTGAACA-CGTACAGGAAATGGAACAAAT---GAAAGAGTCTTT 2695
|||

Db 921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 941
|||
Qy 2696 GTGTACTGAAAGAAC---TGTCAACAACAAAGAAATAAATCAAGTAGAGAA---CCA 2749
|||
Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluLeuLysSerGln-LeuGluAsnGln 960
|||
Qy 2750 AAAGTAAATGGAAACAAGAG---TCTCAGTGTAGATTGACTTAAACCAA---CAAAAG 2803
|||
Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
|||
Qy 2804 AGAAGAGAAATCCGATATATTAAAGAAAAA-----TTAGAGAAGATTAG 2851
|||
Db 981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1000
|||
Qy 2852 AAGATTCGAGACACATAGAAAGATTAGAGTGAACACACAACTTGAAG---GCTCTC 2908
|||
Db 1000 LuGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
|||
Qy 2909 AGATACAAG---ATAAGAATTGAAGTGTAGAAAGTAAATTTCAATAGGTTTCTCA---AC 2962
|||
Db 1018 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
|||
Qy 2963 TCATGAATGAAA---ATTATCTTACATGAATTCGATGTTGAAAGAGGAATTGCCA-- 3017
|||
Db 1037 HisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
|||
Qy 3018 -TGCAAAACTGGAATAGCCACACTA---AACACCAATCCAGAAAGAAAGAAATAAATC-T 3072
|||
Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
|||
Qy 3073 TTGAGACTTAAGATTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
|||
Db 1077 heGluAspIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1097
|||
Qy 3129 AAAGGAATCATTACTATAAG---CGATCTCAATATGTGGGCAGCT---AAAGTCTCATA 3182
|||
Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
|||
Qy 3183 GCTGAGAAACAATGCTCATCTTAATTTGAGGAAACAGACAAAGAAAA-----CTAGA 3236
|||
Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1136
|||
Qy 3237 GGCAGAAATTCGAATA-CACCATCTGCTGCTCGCTGTGTACAGC-----CA 3283
|||
Db 1136 uAlaGluLeuSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
|||
Qy 3284 TGATCAATGTGACATCAAGAAAAAGTAAGA---ACCTGCTTCCACATTGAG---GAGAT 3337
|||
Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
|||
Qy 3338 GCTGTTTCAAGAAATGAATGTTGATGTG---AGAGTACGATATTAAACATGAGTGCT 3394
|||
Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerThrIleTyr-AsnAsnGluValL 1193
|||
Qy 3395 CCATCACCACCTTTC---TGAGCTCAAAAGGAAATCCAAACCTCAAAATTTATCT---CAAT 3448
|||
Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
|||
Qy 3449 TAGCAGGAGATGCTTAAGAGA---AATACATGTTTTCAGACATCATCAAGAGA---CA 3502
|||
Db 1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
|||
Qy 3503 ACGTGAACACAGTGTCAAT---GAAGGAAGCGCAACACATGTATCAAAAC-----GACAA 3553
|||
Db 1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
|||
Qy 3554 GATTAATGGAACAAACACCTGACAGCAGAGTCTCTAGTACAGATTTATTTCACTACAAG 3613
|||
Db 1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGln 1272
|||
Qy 3614 CAAAAATTTGGCTTTCACGCAATT---AGTCATGCACATAGAAAAGCTGACACAAAG 3669
|||
Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
|||

QY 3670 CAGATAACAAT-----GATATTCAATTTCTTG---AGAGCAAAATGCAACAC 3711
 Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluLargLysMetGlnHisH 1308
 QY 3712 ATCTCTAAAGAGAAAAAGAGAGATATT---AATTACAATACCATTTTAAAAACCGTAT 3768
 Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
 QY 3769 ATT---CAATATCAAAAAGAGAAA 3789
 Db 1328 IleTyrGlnTyrGlnLysGluLys 1335

RESULT 5
 ABJ37788
 ID ABJ37788 standard; protein; 1349 AA.
 XX AC ABJ37788;
 DT 15-MAY-2003 (first entry)
 XX Human tumour-related protein - SEQ ID NO 573.
 DE Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
 KW tumour; breast cancer; cancer; immune response stimulation.
 XX OS Homo sapiens.
 XX WO200283956-A1.
 XX 24-OCT-2002.
 XX 15-APR-2002; 2002WO-US012378.
 XX 13-APR-2001; 2001US-00834759.
 PR 07-DEC-2001; 2001US-00007805.
 PR 13-FEB-2002; 2002US-00076622.
 XX (CORI-) CORIXA CORP.
 PA Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
 PI Vedvick TS, McNeill PD, Durham M;
 XX WPI; 2003-103376/09.
 XX New polypeptide and polynucleotide useful for stimulating and/or
 PT expanding T cells specific for a tumor protein and treating breast
 PT cancer.
 XX Example 12; Page 353-357; 375pp; English.
 PS
 CC The invention comprises a method of stimulating and/or expanding T cells
 CC specific for a tumor protein. The invention further comprises human
 CC nucleic acids and proteins that are associated with tumors (e.g. breast
 CC cancer). The method and sequences of the invention are useful for
 CC stimulating and/or expanding T cells specific for a tumor protein,
 CC detecting the presence of cancer, stimulating an immune response in a
 CC patient and treating breast cancer. The present amino acid sequence
 CC represents a human tumour-related protein
 XX
 SQ Sequence 1349 AA;

Alignment Scores:
 Pred. No.: 1,92e-130 Length: 1349
 Score: 1223.00 Matches: 677
 Percent Similarity: 59.07% Conservative: 150
 Best Local Similarity: 48.36% Mismatches: 380
 Query Match: 17.26% Indels: 207
 DB: 6 Gaps: 115

US-09-602-362E-22 (1-4115) x ABJ37788 (1-1349)

QY 96 CAAGAAGGAAGAAGACATCAACCTTTAATATACAAGAG---CCCAGAAGAGAGCTGCTCTAAC 152
 Db 11 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 29
 QY 153 TGGGCCTGGTCAATGGCCT-----GAGAAAGTAGTAACATTTC---CGTA 194
 Db 30 -----HisrPalCysValAsnGlyHisGluGluValValThr-PheLeuVala 46
 QY 195 GACAAAGTGCACCTGACGTCCCTTG---TGGCGAAACACGACACACCTCTGATGAAGGC-T 250
 Db 46 sPArgLysCysGlnProaspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 66
 QY 251 TACAATGCCACAGAGGCTTTCG---AAATATTGATAGATTC---GGTCCCGATATAATC 306
 Db 66 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 86
 QY 307 TCGTG---ATGTGTATGCAACATGGCTTCCATTATGCGTTTATAGTGATTTTCTCAGT- 362
 Db 86 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluLeuLeuSerV 106
 QY 363 --GGTGGCAAACTGCTGTC-CATGGTCAGTATCGAAGTGC---CAACAAGGCTGCCTCA 416
 Db 106 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 126
 QY 417 CACCACCTTTTACTACCAT---AACGAAAGAGTGCAGCAATTGTGGAATTTT---GCTGAT 470
 Db 126 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuLe 145
 QY 471 AAAAATGCAAAATGCGAATGCGTTA---ATAAGTTAAATGCACACCTCATGCTCTGTAT 527
 Db 146 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 165
 QY 528 G---TCTGGATCATCAGAGATAGTTGCATGCTTCTTAGCAAAA---TGTGACGCTTTT-- 579
 Db 165 ICysHisGlyLeuSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 185
 QY 580 -GTGCAGATATATGGAGTAACCTGCAGAACAT---ATGCTCTTACTGCTGGATTCTTCACA 635
 Db 185 aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisI 205
 QY 636 TTCATAACAAA---TTATGAATATATACGAAAAATATCAAAAATCA---TCAATATACAA 689
 Db 205 leHisGluGlnIleMet-GluTyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 224
 QY 690 T---CAGAAGGAACCTCTGCAGAAAACCTGATGAGGCTGCACCTC---GGCGAAAGACAC 743
 Db 224 nProGluGlyThrSerAlaGlyThrProaspGluAlaAla-ProLeuAlaGluArgThrP 244
 QY 744 CTGACACGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGCTGCACCTC---GGTGGGA 797
 Db 244 roAspThrAlaGluSerLeuValGluLysThrProaspGluAlaAla-ProLeuValGlu 263
 QY 798 AAGACACCTGCAGGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGCTGCATCT--- 851
 Db 264 ArgThrProaspThrAlaGluSerLeuValGluLysThrProaspGluAlaAlaSerLeu 283
 QY 852 TGGTGGAGGACATCTGCACAAATTCATTTGGAGAAGG---ACATCTGAAAAGTTTCTGA 908
 Db 284 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 303
 QY 909 AGCTCAGC---AGAAAAACACCTAGGAATACATCC---TGCAAAAAACATCTCAGA 962
 Db 303 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 323
 QY 963 AATTTTACG---GGCCAGCAAAAGGAAGACCTAGAAGATCGCAGGAGAGAA---AAAGAAGAC 1016
 Db 323 ySPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluasp 342
 QY 1017 ACACCTAGGAAATATAGTCCCGAAAAGAAACACTGAGAAA-----TTA 1061
 Db 343 ThrProArgGluIleMetSerProAlaLysGluThrSerGlyLysPheThrTrpAlaAla 362
 QY 1062 CGTGGCAGAAAAGGAAGACACCTAGGAAGCCATGGGAGAAAAGAAA---CCCTGTAAA 1118

	Qy	3073	TTGAGGACTTAAGATTITAAAGA---AAAGATGTGAACCTTAG-ATGACCCTAAAACCTGA	3128
	Db	1085	heGlusAspIleLeuLysLeuLysSerAlaGluLeuGlnMetThrLeuLysLeuL	1105
	Qy	3129	AAAGGAATCATTACTATAAAG---GGATCTCAATATGTGGGCAGCT---AAAGTTCCTCATA	3182
	Db	1105	ySGluGluSerLeuThrLyserArgAlaserGlnTySerGly-GlnLeuLysValLeuile	1124
	Qy	3183	GCTCAGAGAACNATGCTCATCTCTAAATTCAGGAAAAACAAGACAAGAAAA-----CTAGA	3236
	Db	1125	AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluileLeuGL	1144
	Qy	3237	GGCAGAAATTCGAATA-CACCATCCCTGACTGCCCTCGCTGTACAAGC-----CA	3283
	Db	1144	uAlaGluilegluSerHisshisPro-----ArgLeuAlaserAlavalGlnaspHi	1161
	Qy	3284	TGATCAATTTTGACATCATCAAGAAAAAGTAAGA---ACCTGCTTCACATTAG--GAGAT	3337
	Db	1161	sasPGinIleValThrSerArgLysSerGlnGluProAlaPheHisiIealagLysAspal	1181
	Qy	3338	GCTGTTTC AAAAGAAAATGAATGTGATGTG---AGAGTACGATATTAAACAATGAGTGCT	3394
	Db	1181	aCyseuLeuGlnargLysMetasnValaspValSerSerThrIleTyrr-AsnasngluVall	1201
	Qy	3395	CCATCACCATTTTC---TGAGCTCAAAGGAAATCCAAAACCTAAAATTATCT---CAAT	3448
	Db	1201	euihsGlnProLeuSetrglualagnlarginylsserLysSerLeuLysIleasnLeuasnt	1321
	Qy	3449	TAGCAGAGATGCTTAAGAGA---AATACATTCGGTTTCACAACATCAACAAGAGA---CA	3502
	Db	1221	yrAlaGlyAsp-Alaleuar-glu-asnthrLeuvalserGluhis-AlaGlnargaspGI	1240
	Qy	3503	ACGTGAANAACAGTGTCNAAT---GAAGGAAGCAACACATGATCAAAAAC-----GACMA	3553
	Db	1240	na-sgLuThr-GlnCysGlnMetLysGluAlaGluHisMetTyrr-GlnAsnGluInaspas	1260
	Qy	3554	GATAATGGRACAAACACCTCAACAGCAGAGCTCTAGTCAGAAATTTTCACTACAAG	3613
	Db	1260	nvalasnLysHishiThrGluGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns	1280
	Qy	3614	CAAAAAATTGTGGCTTCACAGCAATT----AGTCATGCACATAGAAAGCTGACAAACAAAAG	3669
	Db	1280	erLysasnMetTrpLeuGlnGlnLeuValHisalahisLys-----LYsa	1296
	Qy	3670	CAGATAACAAT-----GATATTCATTCTTTG---AGAGGAAATGCAACAC	3711
	Db	1296	laAspasnLysSerLysIleThrIleAspileHisPheLeuGluArGlysMetGlnHis	1316
	Qy	3712	ATTCCTCTAAAGAGAAAAAGCAGGAGATATT---AATTACAATACCATTAAAAACCGTAT	3768
	Db	1316	islLeuleuLysGluLysAsnGlu-GluillePheashnTyraSnAsnHisiLeuLysAsnArg	1335
	Qy	3769	ATT---CAATATGAAAAAGAGAAGAAACAGAAAACACTCATG	3807
	Db	1336	IleTyrrGlnTyrrGluLysGluLysAlaGluThrGluValile	1349

RESULT 6

RESULT 6
ABJT37783

AB037783
ID ABJ37783 standard: protein: 1013 AA.

XX
XX
COPY 57

AC ABJ37783;

XX

DT 15-MAY-2003 (first entry)

XX

DE Human tumour-related protein - SEQ ID No 553.

XX

Human; vaccine; gene therapy; T cell stimulation

KW
yy
tumour; breast cancer; cancer; immune respons

CC-0
CC-1
CC-2
CC-3
CC-4
CC-5
CC-6
CC-7
CC-8
CC-9

XX	24-OCT-2002.	
PD		
XX		
XX	15-APR-2002; 2002WO-US012378.	
PF		
XX		
XX	13-APR-2001; 2001US-00834759.	
PR	07-DEC-2001; 2001US-00007805.	
PR	13-FEB-2002; 2002US-00076622.	
XX		
XX	(CORI-) CORIXA CORP.	
PA		
XX		
XX	Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;	
PI	Micham JL, Xu J, Harlocker SL, Hepier WT, Henderson RA, Fanger GR;	
PI	Wedvick TS, McNeill PD, Durham M;	
XX		
XX	WPI; 2003-103376/09.	
DR		
XX		
PT	New polypeptide and polynucleotide useful for stimulating and/or	
PT	expanding T cells specific for a tumor protein and treating breast	
PT	cancer.	
XX		
XX	Example 8; Page 335-338; 375pp; English.	
PS		
XX		
CC	The invention comprises a method of stimulating and/or expanding T cells	
CC	specific for a tumour protein. The invention further comprises human	
CC	nucleic acids and proteins that are associated with tumours (e.g. breast	
CC	cancer). The method and sequences of the invention are useful for	
CC	stimulating and/or expanding T cells specific for a tumour protein,	
CC	detecting the presence of cancer, stimulating an immune response in a	
CC	patient and treating breast cancer. The present amino acid sequence	
CC	represents a human tumour-related protein	
XX		
XX	Sequence 1013 AA;	
SQ		

PN WO200283956-A1.

Db 197 spLysIleAsnGlyLysLeuGluSerProAsnLysAspGlyLeuLysAlaThrC 217
QY 1631 CGGAAT---GAAATTTCTATTCCACTAAGCCT---AGAATTGAAGACATGCAAACTTTC 1684
Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThrPhe 236
QY 1685 AAAGC-GAGCCTCCGGGAAGC---CATCTGCTTCGAGCCTCCACTGAATC---AAAAGT 1737
Db 237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
QY 1738 CTGTCCCAATATAGCCTTGGAAATGAAATGAAACATGAGGCAGATGAGAA---CTC 1794
Db 257 ValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGluLysLeu 276
QY 1795 CCATCAGAAATCCAAACAAAGGACTATAAGAAA---TTCTGGATCTGAGTCTCTG--- 1849
Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSer-GluSerLeuCy 296
QY 1850 -TAGACTGTTTCAGAGGATTC-TGTTTACCAAGG---CTGCGCTCAAAAAGAAATAGA 1904
Db 296 sGluThr-ValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluLys 316
QY 1905 TAAATATGAAATATAGAGGT---CCCTGTTAAAGTGGTCTCTGAGGCTAACT-- 1959
Db 316 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLysAlaAsnCy 336
QY 1960 -GGAAATGAAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACATGCAAACTTCAA 2015
Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLysMetAspMetGlnThrPhe 356
QY 2016 AG---CAGGCTCCGAGAGCCTCTGCTTCGAGCCTCCATTGAAA---TCAAAAGTCT 2069
Db 356 ysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
QY 2070 GTCCAAATAAAC---CTTGGAAATTAAGATGAACAAACATGAGGCAGATGAGAA---CT 2123
Db 376 alProAsn-LysAlaLeuGluLysAsnGluLysAsnGluLysAlaAsp-GluLysLe 395
QY 2124 CCATCA---AATCCAAACAAAGGACTATGAGAAAGTTCTTGGAGTCTCTCTG- 2179
Db 395 uProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSer-GluSerLeuC 415
QY 2180 --TAGACTGTTCCAGAGATCGTG---TTTACCCAGCCTACACATCAAAAGAAATAGA 2234
Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluLysLe 435
QY 2235 TAAAA---TAATGGAATAAGAGAGTCTCT-GATAATGAGGTTTCTGAGGCTCCCT- 2289
Db 435 spLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProC 455
QY 2290 --CAGAAATGAAAGTTCTATTCCACTAAGCCT---TGAATTGATGCGATGCAAACTTCA 2344
Db 455 ysArgMetLysValSerIleProThrLysAlaLeuGluLysMetAspMetGlnThrPheL 475
QY 2345 AAGCAGGCTCCCG---AGAGCCATCTGCTTCGAGCCTGCATTGAAAT---GAAAAGTCT 2398
Db 475 ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
QY 2399 GTTCCAAATAAC---CTTGGAAATTAAGATGAACAAACATGAGCAGATCAGAT- 2450
Db 495 ValProAsn-LysAlaLeuGluLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
QY 2451 -TTCCCTTCAGAAATCAAAACAAAGAGTTGAGAAA---TTCTGGATTTCTGAGTCT- 2505
Db 513 tPheProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGlu-SerL 533
QY 2506 --CCGTGGAGTCTTTACAGAGAGTGGTGTGTACC---CAGGCTACACAAAAGAAA 2560
Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
QY 2561 GCATAAAAT---AGTGGAAATTAAGATTCACCTAGCCTATCAAAAT---CTTGGTACA 2614

Db 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
QY 2615 GTTCAT---CTTGTGAAAGCGAAGGAACTTCAAAAAGACACTGTGAACA-CGTACAGGA 2670
Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
QY 2671 AAATGGACAAT---GAAAAGAGTGTGTGTACTGAAAAAGAAC---TGTCAGAACAAA 2724
Db 593 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuSer-GluAlaLys 612
QY 2725 AGAAATAAATCACAGTACAGAA---CCAAAAGTTAAATGGGAACAAGAG---TCTGCAGT 2778
Db 613 GluLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGluLeuLysSerVa 632
QY 2779 GTAGATTGACTTAAACCAA---GAAAAGAGAGAGAAATGCCGATATATAAAGAAA 2835
Db 632 LArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGluL 652
QY 2836 AA-----TTAGGAAGAATTAGAAGAATCGAGAGCAGCATAGAAAGTTAGAA 2883
Db 652 ysIleArgGluGluLysGlyArgIle-GluGlnGlnHisArg-----LysGluLysGlu 669
QY 2884 GTGAAACAACAATTGAAAG---GCTCTCAGATACAAG---ATAAGAATTGAAGTGTAGA 2937
Db 670 ValLysGlnGlnLeu-LeuGluGlnAlaLeuArgIleGlnAspIleGluLysSerValG 689
QY 2938 AAGTAATTTGAATAGGTTTCTCA---ACTCATGAATGAAA---ATTATTCTTACATGAA 2991
Db 689 uSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLysHisGlu 709
QY 2992 ATTGCATGTTGAAAAAGGAATTGCCA---TGCAAAACCTGGAATAGCCACACTA---AACA 3045
Db 709 snCysMetLysLysGluLysGluLysLeuAlaMetLysLysGluLysAlaThrLeuLysHis 728
QY 3046 CCAATCCAGNAAGGAAATAATC-TTTGAGGACTTAAGATTTAAAGA---AAGAT 3101
Db 729 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysLysLeuLysGluLysAsn 748
QY 3102 GCTGAACTTAG-ATGACCTTAAACTGAAAAGGAATCATTACTATAAAG---GGATCTCAA 3157
Db 749 AlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGln 768
QY 3158 TATGTGGCAGCT---AAGTTCTGTAGTCTGAGAAACAATGCTCATCTTAAATAGGAGA 3214
Db 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 788
QY 3215 AAAACACAGCAAAAGAAA-----CTAGAGCAGCAAAATTTGAATA-CACCATCTGACTGGC 3267
Db 788 ysGluLysGlnAspLysGluLysGluLysGluLysGluLysGluSerHisHisPro----- 805
QY 3268 TTGCTCTACAAGC-----CATGATCAATTGTGACATCAAGAAAAAGTAAAGA- 3314
Db 806 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 825
QY 3315 --ACCTGCTTCCACATTGAG---GAGATGCTGTTTCCAAAGAAAATGAATGTTGATGTCG- 3368
Db 825 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspVal 845
QY 3369 --AGAGTACGATATTAAATAGTGTCTCCATCACCACCTTTC---TGAGCTCAAGAGAAA 3423
Db 845 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
QY 3424 TCCAAAACCTTAAATATATCT---CAATTAGCAGGAGATGCTTAAGAGA---AATACATT 3477
Db 865 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 884
QY 3478 GGTTCAGACATCATCAAAAGAGA---CAACGTGAAACACAGTGTCAAT---GAAAGAGCG 3531
Db 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904
QY 3532 AACACATGATCAAAAC-----GACAAGATATGAAACAAACACCTGAAACAGCAGATC 3585
Db 904 LuHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnSerL 924

3586 TCTAGTCAGAAATATTTCACCTACAAAGCAAAATTCGTGCTTCACAGCAATTT---AGT 3641
 Db eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 943
 3642 CATGACATAGAAAGCTGACACAAAGAGCAGATACAAAT-----GATATT 3686
 Db HisAlaHisLys-----LysAlaAspAsnLysSerLysLeuThrLeuAspIle 959
 3687 CATTTCTTG---AGAGGAAATGCAACACATCTCTCTAAAGAGAAAGAGGAGATATT-- 3741
 Db HisPheLeuGluArgLysMetGlnHisHisLeuLeuLysGluLysAsnGlu-GluIlePhe 979
 3742 -AATTACATACCATTTAAACCGTATATT---CAATATGAAAGAGAGAAA 3789
 Db eAsnTrpAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGluLysGluLys 996

RESULT 8
 ID ABG78918 standard; protein; 1002 AA.
 AC ABG78918;
 XX
 XX Human breast tumour polypeptide #10.
 DT 15-NOV-2002 (first entry)
 XX Human; breast tumour protein; breast cancer; cytostatic; vaccine.
 XX Homo sapiens.
 OS
 XX US2002085998-A1.
 FN
 XX 04-JUL-2002.
 PD
 XX 13-APR-2001; 2001US-00834759.
 XX 28-DEC-1998; 98US-00222575.
 PR 02-APR-1999; 99US-00285480.
 PR 23-JUN-1999; 99US-00339338.
 PR 03-SEP-1999; 99US-00389681.
 PR 17-APR-2000; 2000US-00433826.
 PR 17-APR-2000; 2000US-00551621.
 PR 08-JUN-2000; 2000US-00590751.
 PR 22-JUN-2000; 2000US-00604287.
 PR 20-JUL-2000; 2000US-00620405.
 XX (CORI-) CORIXA CORP.
 FA
 XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepier WT;
 PI Henderson RA;
 XX WPI: 2002-635657/68.
 DR N-PSDB; ABS64012.
 XX
 PT Novel breast cancer polynucleotides and polypeptides encoded by the
 PT polynucleotides, useful for detecting the presence of breast cancer
 PT patient, and in pharmaceutical compositions, for treating breast cancer.
 XX
 PS Claim 2; Page 208-211; 247pp; English.
 XX
 CC The invention relates to an isolated breast tumour polynucleotide and the
 CC polypeptide it encodes. The polynucleotide and polypeptide are useful for
 CC detecting the presence of breast cancer in a patient, and in
 CC pharmaceutical compositions for treating breast cancer. The sequences are
 CC useful for stimulating an immune response in a patient and can therefore
 CC be used in production of vaccines. The sequences are also useful for
 CC detecting the presence of a cancer in a patient by obtaining a
 CC biological sample from the patient, contacting the biological sample with
 CC a composition of the invention and detecting the amount of polynucleotide
 CC that hybridizes to the sample. This sequence represents a human breast
 CC tumour polypeptide of the invention

SQ Sequence 1002 AA;
 Alignment Scores:
 Pred. No.: 2,85e-95 Length: 1002
 Score: 917.00 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.94% Indels: 145
 DB: 5 Gaps: 83

US-09-602-362E-22 (1-4115) x ABG78918 (1-1002)

QY 1031 ATGAGTCCCGAAAGAAACACTGAGAA--ATTACCTGGCAGCA-AAAGGAAGACCTAGG 1086
 Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProCys 20
 QY 1087 AAGAC-CCATGGGAGAGAAAGAAA---CCCTGTAAAGATGGATGCGCT-GCAAGAGTAACA 1141
 Db LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
 QY 1142 TCTAATAAATAAG---TTTGTAAAAGAAATCTAAGATGTT-GCATGTCCACAAAG 1197
 Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
 QY 1198 ---AATCATCTACAAAGCAAGTGCATGATCAGAGG---TCCCATCAGATCCAAACAAGG 1251
 Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
 QY 1252 GAAGATCAAGAATATCTTT-GATTCTCGGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
 Db 81 GluAspGluTyrSerCysAspSerArg--SerLeuPheGluSerSerAlaLysIle 99
 QY 1305 CAAGTG---GTATACCTGAGTCTATATAAAAGTAATCAGATAAATAAGAG---TAG 1358
 Db 100 GlnValCysIleProGluSerIleTyrGlnLysValMetGluLeuAsnArgGluValGlu 119
 QY 1359 AAGCCCTCTAAAGCATCTGCTTCAAGCCGCATTTAAAGCAAAA---CTCTTTCCA 1415
 Db 120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
 QY 1416 AATAAGCCTTTG-----AATGAAGAAATGAACAAACATTTGAAGCAGATCCGTG-TTC 1465
 Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspMetPhe 157
 QY 1466 CCACCG---AATCCAAACAAAGGACTATAGAAAATCTTGGATTTCTAGATCTCTG-- 1520
 Db 158 ProGluSerLysLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
 QY 1521 -GAGACTGTTTACAGAGG---ATTGTGTTTACCAGGCTACACATCAAAAAGAAATAGAT 1576
 Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
 QY 1577 AAAAA---AATGAAAA---TAGAAGATCCCTAATAAGTGGTCTTCTGAAGGCTACT 1630
 Db 197 sPlyIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
 QY 1631 CGGAAT---GAAATTTCTATTCACCTAAAGCCT---AGAATTGAAGACATGCAAACTTTC 1684
 Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
 QY 1685 AAGCC-GAGCCTCCGGGAAGC---CATCTGTTCCAGCCTCCCACTAAATC---AAAAGT 1737
 Db 237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
 QY 1738 CTCTCCAAATAGCCTTGGATGAAAATGAAAACATGAGGAGGAGATGAGAA---CTC 1794
 Db 257 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 276
 QY 1795 CMTAGATCCAAACAAAGGACTATAGAAA---TTCTGGGATCTGGAGTCTCTG-- 1849
 Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 296
 QY 1850 -TAGACTGTTTACAGAGGAGATTG-TGTTTACCAAGG---CTGCGCTCAAAAAGAAATAGA 1904

Db 296 sgluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleAs 316
 QY 1905 TAAATAATCGAAATAGAGGCT---CCCTGTTAAAGTGGTCTCTGAGGCTAACT-- 1959
 Db 316 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 336
 QY 1960 -GGGAATGAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACATCAAACTTCAA 2015
 Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
 QY 2016 AG---CAGGCTCCGAGAGCCATCTCCCTCGAGCTCCATTTGAAA---TCAAAAGTCT 2069
 Db 356 ysaGluThrProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
 QY 2070 GTCCAAATAAAC---CTTGAATTAAGATGAACAACATTGAGGCAGATGAGAA---CT 2123
 Db 376 alProAsn-LysAlaLeuLysAsnGlnThrLeuArgAlaAsp-GluIleLe 395
 QY 2124 CCATCA---AATCCAAACAAGAGCTATGAGAAAGTTCTGGATCTCGATCTCTG- 2179
 Db 395 uProSerGluSerLysGlnLysAspTyrGluGluSerSerTyrAspSer-GluSerLeuC 415
 QY 2180 --TAGACTGTTCCAGAGGATGGTG---TTTACCAGGCTACACATCAAAAGATAGA 2234
 Db 415 ysaGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleA 435
 QY 2235 TAAAA---TAATGAAATAGAGAGTCTCT-GATAATGAGGTTTCTGAAGCTCCCT- 2289
 Db 435 sPlyIleAsnGlyLysLeuGluGlySerProAspAsnAspGlyPheLeuLysAlaProC 455
 QY 2290 --GAGATGAAGTTCTATTCACCTAAAGCCT---TGAATTGATGCATGCAAACTTCA 2344
 Db 455 ysaGlyMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
 QY 2345 AAGCAGGCTCCCG---AGAGCCATCTGCTTCAGGCTCGCATTAAT---GAAAAGTCT 2398
 Db 475 ysaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
 QY 2399 GTTCCAAATAAAC---CTTGAATTAAGATGAACAACATTGAGAGATCAGAT- 2450
 Db 495 ValProAsn-LysAlaLeuLysLeuLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
 QY 2451 -TTCCCTTCAGATCAAAACAAGAGTTTGAGAAA---TTCTGGATTCTGAAAGTCT- 2505
 Db 513 tPheProSerGluSerLysGlnLys**ValGluGluAsnSerTyrAspSerGlu-SerL 533
 QY 2506 --CCGTGGAGTGTTCACAGAGGATGGTGTGTTACC---CAGGTACACAAAAAGAAA 2560
 Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
 QY 2561 CGATAAAAT---AGTGGAAATTAAGAGATTCTACTAGCCATCAAAAT---CTTGGTACA 2614
 Db 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLysLysIleLeuAspThr 572
 QY 2615 GTTCAT---CTTGTGAAGCGCAAGGAAGTTCAAAAGAGACACTGTGAACA-CGTACAGGA 2670
 Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
 QY 2671 AAATGGAACAAT---GAAAAGAGTTTGTGTACTGAAAAGAAC---TGTACAGACAAA 2724
 Db 593 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLysLysLysLysLys 612
 QY 2725 AGAAATAATCACAGTAGAGAA---CCAAAAGTTAAATGGGAACAAGAG---TCTGCACT 2778
 Db 613 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnLeuLysSerVa 632
 QY 2779 GTAGATTGCTTAACCAA---GAAAAGAGAGAGAGAAATGCCGATATATAAGAAAA 2835
 Db 632 lArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGluL 652
 QY 2836 AA-----TTAGAGAAATTAGAAGATCGAGAGCAGCATAGAAAGAGTTAGAA 2883

Db 652 ysaIleArgGluLeuLysArgIle-GluGlnGlnHisArg-----LysGluLeuGlu 669
 QY 2884 GTGAACAACAACCTTGAAG---GCTCTCAGATACAG---ATAAGAATTGAAGGTGTAGA 2937
 Db 670 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValG 689
 QY 2938 AAGTAATTGATAGTTTCTCA---ACTCATGAATGAAA---ATTATCTTTCATGAA 2991
 Db 689 userAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGlu 709
 QY 2992 ATTGATGTTGAAAAGAAATTGCCA---TGCAAAATCGAATAGCCACACTA---AACA 3045
 Db 709 snCysMetLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis- 728
 QY 3046 CCAATCCAGGAAAAGAAATAAATC-TTTGAGGACTTAAGATTTAAAGA---AAAGAT 3101
 Db 729 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsn 748
 QY 3102 GCTGAACCTTAG-ATGACCTTAAACTGAAAAGGAATCATTTACTATAAAG---GGAATCTCAA 3157
 Db 749 AlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGln 768
 QY 3158 TATGTGGCAGCT---AAAGTCTCTGATAGCTGAGAAAACAATGCTCATCTTAATTTAGGA 3214
 Db 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 788
 QY 3215 AAAACAGACAAAAGAAA---CTAGAGGCGAGAAATTTGAATA-CACCATCTCTGACTGGC 3267
 Db 788 ysaGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisPro----- 805
 QY 3268 TTCCTGTACAGC-----CATGATCAATTTGTGCATCAAGAAGAAAGTAAAG- 3314
 Db 806 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 825
 QY 3315 -ACCTGCTCCACATTTAG---GAGATGCTGTTTGCAAAAGAAAATGAATGTTGATGTG- 3368
 Db 825 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 845
 QY 3369 --AGATGACGATATTAACAATGATGCTCCATCCACCATTTTC---TCAGCTCAAGGAAA 3423
 Db 845 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
 QY 3424 TCCAAAACCTTAAATTTATCT---CAATTAGCAGGAGATGCTTAAGAGA---AATACATT 3477
 Db 865 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 884
 QY 3478 GGTTCAGAACATCAACAAGAGA---CAACGTGAAAACACAGTGTCAAT---GAAGGAAGCG 3531
 Db 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904
 QY 3532 AACACATGATCAAAAC-----GACAAGATATGAAACAACACCTGACACAGCAGATC 3585
 Db 904 luHisMetTyrGlnAsnGluGlnAspValAsnLysHisThrGluGlnGlnSerL 924
 QY 3586 TCTAGTCAGAAATTTATTCAACTACAGCAAAAATTTGGCTTTCACAGCAAT---AGT 3641
 Db 924 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 943
 QY 3642 CATGCACATAGAAGCTGACACAAGAGCAGATAACAAT-----GATATT 3686
 Db 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
 QY 3687 CATTTCTTG---AGAGGAAAATGCAACACATCTCTCTAAAAGAGAAAAGAGGATATT-- 3741
 Db 960 HisPheLeuGluArgLysMetGlnHisLeuLeuLysGluLysAsnGlu-GluIlePh 979
 QY 3742 -AATTCATACATTATAAAACCGTATATT---CAATATGAAAAGAGAAA 3789
 Db 979 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGluLysGluLys 996
 RESULT 9
 ABJ37741
 ID ABJ37741 standard; protein; 1002 AA.

XX AC ABJ37741;
 XX DI 15-MAY-2003 (first entry)
 XX DE Human tumour-related protein - SEQ ID No 475.
 XX KW Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
 XX KW tumour; breast cancer; cancer; immune response stimulation.
 XX OS Homo sapiens.
 XX FN WO200283956-A1.
 XX PD 24-OCT-2002.
 XX PF 15-APR-2002; 2002WO-US012378.
 XX PR 13-APR-2001; 2001US-00834759.
 XX PR 07-DEC-2001; 2001US-00007805.
 XX PR 13-FEB-2002; 2002US-00076622.
 XX PA (CORI-) CORIXA CORP.
 XX PI Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 XX PI Mitcham JL, Xu J, Hazlocke SL, Hepler WT, Henderson RA, Fanger GR;
 XX PI Vedvick TS, McNeill PD, Durham M;
 XX DR WPI; 2003-103376/09.
 XX PT New polypeptide and polynucleotide useful for stimulating and/or
 XX PT expanding T cells specific for a tumor protein and treating breast
 XX PT cancer.
 XX PS Example 1; Page 298-300; 375pp; English.
 XX CC The invention comprises a method of stimulating and/or expanding T cells
 XX CC specific for a tumour protein. The invention further comprises human
 XX CC nucleic acids and proteins that are associated with tumours (e.g. breast
 XX CC cancer). The method and sequences of the invention are useful for
 XX CC stimulating and/or expanding T cells specific for a tumour protein,
 XX CC detecting the presence of cancer, stimulating an immune response in a
 XX CC patient and treating breast cancer. The present amino acid sequence
 XX CC represents a human tumour-related protein
 XX SQ Sequence 1002 AA;

Alignment Scores:
 Pred. No.: 2,85e-95 Length: 1002
 Score: 927.00 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.94% Indels: 145
 DB: 6 Gaps: 83

US-09-602-362E-22 (1-4115) x ABJ37741 (1-1002)

QY 1031 ATGAGTCCGMAAAGAAACACTGAGAA---ATTACGTGGCAGC-AAAGCAGACCTAGG 1086
 Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 20
 QY 1087 AAGAC-GCATGGGAGAAAGAA---CCCTGTAAGATGGATGCGT-GCAAGAGTAACA 1141
 Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
 QY 1142 TCTAATTAACATAAG---TTTTGAAAGGAATCTAAGATGTT-CCATGTCCACAAAG 1197
 Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
 QY 1198 ---AATCATCTACAAAGCAAGTCCATGATCAGAGG---TCCCATCAGATCCAAAG 1251
 Db 51 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80

QY 1252 GAGATGAGAATATTTT-GATTCCTGGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
 Db 81 GluAspGluGlySerCysAspSerArg---SerLeuPheGluSerSerAlaLysIle 99
 QY 1305 CAAGTG---GTATACCTGAGTCTATATATAAAAGTAATAGATAAATAAGAG---TAG 1358
 Db 100 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 119
 QY 1359 AAAGCCCTCTAAAGGCATCTGCTTCAAGCGCCCATTTGAAAGCAAAA---CTCTTTCCA 1415
 Db 120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
 QY 1416 AATAAGCCTTTG-----AATGAAGATGAACAAACATTGAAGCAGATCCGTG-TTC 1465
 Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
 QY 1466 CCACCG---AATCCAAACAAAGGACATATAGAAAATTTCTTGGGATTCTAGATCTCTG-- 1520
 Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
 QY 1521 -GAGACTGTTTACAGAAG---ATTGTGTTTACCAGGCTACACATCAAAAAGAATAGAT 1576
 Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
 QY 1577 AAAAA---AATGGA---TAGAAGAGTCCCTAATAAAGTGTCTTCTGAGGCTACTCT 1630
 Db 197 spLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
 QY 1631 CGGAAT---GAAATTTCTATTCACATAAGCCCT---AGAATTGAAGACATGCAAACTTTC 1684
 Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThrPhe 236
 QY 1685 AAGGC-GAGCCTCGGGAGC---CATCTGCTTCGAGCCTCCACATAATC---AAAAGT 1737
 Db 237 LysAlaGluProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
 QY 1738 CTGTCCCAATAAGCCTTGGATGAAATGAAACATGGAGCGCAGATGAGAA---CTC 1794
 Db 257 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 276
 QY 1795 CCATCAGAAATCCAAACAAAGGACTATAAGAAA---TTCTGGGATCTGCTCTCTG-- 1849
 Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 296
 QY 1850 -TAGACTGTTTACAGAAGCATTTG-TGTTTACCAAGG---CTGCGCTCAAAAGAAATAGA 1904
 Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleAs 316
 QY 1905 TAAATAATCGAAATAGAGGCT---CCCTGTTAAAGTGGTCTTCTGAGGCTAACT-- 1959
 Db 316 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 336
 QY 1960 -GGGAATGAAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACATGCAAACTTCAA 2015
 Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
 QY 2016 AG---CAGGCCTCCGAGAGCCATCTGCTTCGAGCCTCCATTGAAA---TCAAAGTCT 2069
 Db 356 ysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
 QY 2070 GTCCAAATAAAC---CTTGGAAATTAAGAATGAACAAACATTGAGCGCAGATGAGAA---CT 2123
 Db 376 alProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 395
 QY 2124 CCATCA---AATCCAAACAAAGGACTATGAAAGATCTTGGGATCTTGGAGTCTCTG-- 2179
 Db 395 uProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 415
 QY 2180 --TAGACTGTTTCCAGAGGATGGTG---TTTACCAGGCTACACATCAAAAAGAAATAGA 2234
 Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleA 435
 QY 2235 TAAAA---TAATGAAATAAGAGATCTCT-GATTAATGAGGCTTTTCTGAAAGCTCCCT- 2289

435 spLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProC 455
 2290 --GAGAATGAAGTTCTATTCACATAAAGCCT--TGAATTTGATGGCATGCAAACTTTCA 2344
 455 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
 2345 AAGCAGGCTCCCG--AGAGCCATCTGCTTCGAGCTGATTCGAAT--GAAAAGTCT 2398
 475 ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
 2399 GTTCCAAATAAAC--CTTGAATTAAGAAATGAACAATGACAGATGACAGATCAGAT 2450
 495 ValProasn-LysAlaLeuGluLeuLysAsnGluGlnThr--LeuArgAlaAspGlnMe 513
 2451 -TTCCCTTCAGATCAAAACAAAGAGGTTGAGAAA--TTCTGGATTTCTCAAGTCT- 2505
 513 tPheProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGlu-SerL 533
 2506 --CCGTGGACTGTTTCACAGAGGATGCTGTGTACC--CAGGCTACACAAAGAGAAA 2560
 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
 2561 GGATAAAT--AGTGAATAATAGAGATTCATAGCTATCAAAAT--CTTGTGTACA 2614
 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
 2615 GTTCAT--CTTGTGAAGCAAGGAACTTCAAAAGACACTGTGAAACA-CGTACAGGA 2670
 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
 2671 AAATGGAAACAT--GAAAGAAGATTTGTGTACTGAAAAAGAAC--TGTCAACAAA 2724
 593 LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 612
 2725 AGAAATAATCAGCTAGAGAA--CCTCAAGTTAATGGCAACAGAG--TCTGCAGT 2778
 613 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGluLysCysSerVa 632
 2779 GTAGATTGACTTAAACCAA--GAAAGAGAGAGAGAAATCCGATATATTAAAGAAA 2835
 632 ArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGluL 652
 2836 AA-----TTAGGAAGATTAGAGAAATCGAGACGACATAGAGAGATTAGAA 2883
 652 ysIleArgGluGluLeuGlyArgIle-GluGluGlnHisArg-----LysGluLeuGlu 669
 2884 GTGAAACACAACTTGAAG--GCTCTCAGATACAG--ATAAGAAATGAAGTGTAGA 2937
 670 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValGI 689
 2938 AGTAATTTGAATAGTTTCTCA--ACTCATGAATGAAG--ATTATTCTTACATGAA 2991
 689 uSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrrLeuLeuHisGluA 709
 2992 ATTGCATGTTGAAAGGAATGCCA--TGCAAACTCGAATAGCCACACTA--AACA 3045
 709 snCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis 728
 3046 CCAATCCAGGAAAGAAAATAATC-TTTGAGGACTTAAAGATTTAAAGA--AAAGAT 3101
 729 GlnTyrrGlnLysGluAsnLysTyrrPheGluAspIleLysIleLeuLysGluLysAsn 748
 3102 GCTGAACITAG-ATGACCTTAAACTGAAAGGAATCATTTACTAAAG--GGATCTCAA 3157
 749 AlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGln 768
 3158 TATGTGGGAGCT--AAAGTTCTCATAGCTGAGAAACAATGCTCATTTCTAAATTTGAGA 3214
 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 789
 3215 AAAACAGACAAAGAAA-----CTAGAGGAGAAATGAATA-CACCATCTCACTGGC 3267

788 ysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisPro----- 805
 3268 TTCGTGTACAGC-----CATGATCAATTTGTGACATCAAGAAAAGCTAAGA- 3314
 806 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 825
 3315 --ACCTGCTTCACATTCAG--GAGATGCTGTTTGCACAAAGAAAATGAATCTTGTGTG- 3368
 825 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspVal 845
 3369 --AGAGTACGATTAATAAATGAGTGTCTCCATCACCCTTC--TGAGCTCAAGGAAA 3423
 845 erSerThrIleTyrr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
 3424 TCCAAACCTAAAATTATCT--CAATTAGCAGGAGATGCTTAAGAGA--AATACATT 3477
 865 SerLysSerLeuLysIleAsnLeuAsnTyrrAlaGlyAsp-AlaLeuArgGluAsnThrLe 884
 3478 GGTTCAGAAATCACAAGAGA--CAACGTGAAAACAGTGTCAAT--GAAGGAAGCG 3531
 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904
 3532 AACCATGTATCAAAAC-----CACAAAGATAATGGAACAAACACCTGAACAGCAGATC 3585
 904 luHisMetTyrrGlnAsnGluGlnAspValAsnLysHisThrGluGlnGlnGluSerL 924
 3586 TCTAGTCGAAATTTATTTCACTACAAGCAAAATTTGTGGCTTCACAGCAAT--AGT 3641
 924 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 943
 3642 CATGCACATAGAAAGCTGACAAACAAAGACAGATAACAAT-----GATATT 3686
 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
 3687 CATTTCTTG--AGAGAAATGCAACACATCTCTTAAAGAGAAAAGAGAGATATT-- 3741
 960 HisPheLeuGluArgLysMetGlnHisHisLeuLysGluLysAsnGlu-GluIlePh 979
 3742 -AATTCAATACCATTTTAAACCGTATATT---CAATATGAAAAGAGAAA 3789
 979 eAsnTyrrAsnAsnHisLeuLysAsnArgIleTyrrGlnTyrrGluLysGluLys 996

RESULT 10

ADE4427

ID ADE4427 standard; protein; 1002 AA.

AC ADE4427;

DT 29-JAN-2004 (first entry)

DE Human breast cancer protein #10.

 KW human; breast tumour; cancer; vaccine; T cell stimulator;
 KW T cell expander.

OS Homo sapiens.

XX US2003104366-A1.

XX 05-JUN-2003.

XX 17-APR-2000; 2000US-0051621.

XX 28-DEC-1998; 98US-00222575.

XX 02-APR-1999; 99US-00285480.

XX 23-JUN-1999; 99US-00339338.

XX 02-SEP-1999; 99US-00389681.

XX 03-NOV-1999; 99US-00433826.

XX (JIAN/) JIANG Y.

XX (DILL/) DILLON D C.

XX (MITC/) MITCHAM J L.

XX (XUJG/) XU J.

PA (HARL/) HARLOCKER S L.

PI Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL;

DR WPI; 2004-020270/02.

DR N-PSDB; ADE44426.

XX Novel isolated polypeptide comprising immunogenic portion of breast tumor protein or its variant, useful for formulating vaccines for inhibiting cancer development in a patient.

PS Example 1; SEQ ID NO 475; 217pp; English.

XX The invention relates to an isolated polypeptide comprising at least an immunogenic portion of a breast tumour protein. The polynucleotide, its polypeptide, its antibody, a pharmaceutical composition comprising the fusion protein or the polynucleotide encoding it, a vaccine comprising the fusion protein or the polynucleotide encoding it, an isolated T cell population comprising T cells specific for a breast tumour protein, and a method for removing tumour cells from a biological sample is useful for inhibiting the development of a cancer in a patient. The polypeptide is useful for stimulating and/or expanding T cells specific for a breast tumour protein. Stimulating and/or expanding T cells specific for a breast tumour protein is useful for inhibiting the development of a cancer in a patient. The method additionally involves the step of cloning at least one proliferated cell and then administering the cloned T cells to the patient. The present sequence represents a breast cancer protein.

XX SQ Sequence 1002 AA;

Alignment Scores:

Pred. No.:	2,85e-95	Length:	1002
Score:	917.00	Matches:	506
Percent Similarity:	59.77%	Conservative:	112
Best Local Similarity:	48.94%	Mismatches:	285
Query Match:	12.94%	Indels:	145
DB:	8	Gaps:	83

US-09-602-362E-22 (1-4115) x ADE44427 (1-1002)

QY 1031 ATGAGTCCCGAAGAAACACTAGAA---ATTACGTGGGCGA-NAAGAGACCTAGG 1086
 Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 20
 QY 1087 AAGAC-GCATGGGAGAAAGAA---CCCTGTAAAGATGGATCGCT-CAAGAGTAACA 1141
 Db 21 LysileAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
 QY 1142 TCATTAATACTAAG---TTTGTAAAGAAATCTAAGATGTT-CCATGTCCACAAAG 1197
 Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
 QY 1198 ---AATCATCTACAAGCAAGTCCATGATCAGAGG---TCCCATCAGATCCAAACAAG 1251
 Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
 QY 1252 GAAGATCAAGNATATCTTT-GATTCGGGTCTCTTTG-----AAGTCTGCAAGATT 1304
 Db 81 GluAspGluGluTrpSerCysAspSerArg---SerLysPheGluSerSerAlaLysIle 99
 QY 1305 CAAGTG---GTATACCTCAGTCTATATATAAAGTAAGTAATCAAGATAAAGAG---TAG 1358
 Db 100 GlnValCysIleProGluSerIleTyGlnLysValMetGluIleAsnArgGluValGlu 119
 QY 1359 AAAGCCTCTTAAGACCATCTGCTTCAAGCCGCAATTAAGCAAAA---CTCTTTCGA 1415
 Db 120 GluProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
 QY 1416 AATAAGCCTTTG-----AATGAAGAATGAACAAATTAAGACAGATCCGTG-TTC 1465
 Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspMetPhe 157
 QY 1466 CCACCG---AATCCAAACAAGGACTATAGAATAATCTTGGGATCTAGACTCTCTG--- 1520

Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
 QY 1521 -GAGACTGTTTACAGAAG---ATTGTGTTTACCAAGGCTACACATCAAAAAAGATAGAT 1576
 Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
 QY 1577 AAAAA---AATGAAAA---TAGAAGAGTCCCTAATAAAGTGTCTTCTGAGGCTACCT 1630
 Db 197 spLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
 QY 1631 CGGAAT---GAAATTTCTATTCACCTAAAGCCT---AGAAATTGAAGACATCAAACTTC 1684
 Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
 QY 1685 AAAGC-CAGCTCCGGGAAGC---CATCTGCTTCGAGCTCCACTCAAAATC---AAAAGT 1737
 Db 237 LysAlaGluProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
 QY 1738 CTGTCCCAATAAGCCTTGAATGAATAAATAAACAATGGAGGCGAGATGAGAA---CTC 1794
 Db 257 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 276
 QY 1795 CCATCAATCCAAACAAGGACTATAAGAAA---TTCTGGGATCTCGAGTCTCTG--- 1849
 Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 296
 QY 1850 -TAGACTGTTTCAAGAAGATTG-TGTTTACAAGG---CTGCGCTCAAAAGATAGATA 1904
 Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleAs 316
 QY 1905 TAAATAATGAATAATGAAGGCT---CCCTGTAAAGTGTCTCTCAGGCTAACT--- 1959
 Db 316 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 336
 QY 1960 -GGGAATGAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACATCAAACTTCAA 2015
 Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
 QY 2016 AG---CAGGCTCCCGAGAGCCATCGCTTCGAGCTCCATTGAAA---TCAAAAGTCT 2069
 Db 356 ysaAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
 QY 2070 GTCCAAATAAAC---CTTGAATTAAGATGAACAAACATTCAGGCGAGATGAGAA---CT 2123
 Db 376 alProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 395
 QY 2124 CCATCA---AATCCAAACAAGGACTATCAGAAAGTCTCTGGATTCCTGAGTCTCTG- 2179
 Db 395 uproSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 415
 QY 2180 --TAGACTGTTTCCAGAAGGATGGTG---TTTACCCAGGCTACACATCAAAAGATAGA 2234
 Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleA 435
 QY 2235 TAAAA---TAATGGAATAAGAGTCTCT-GATAATGAGTCTTCTGAGGCTCCCT- 2289
 Db 435 spLysIleAsnGlyLysLeuGluGluSerProAsnAspGlyPheLeuLysAlaProC 455
 QY 2290 --GAGAATGAAGATTCATTCCACTAAAGCCT---TGAATTGATGCGATCAAACTTCA 2344
 Db 455 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
 QY 2345 AAGAGGCTCCCG---AGAGCCATCTGCTCGAGCTCGATGAAAT---GAAAGTCT 2398
 Db 475 ysaAlaGluProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
 QY 2399 GTTCCAAATAAAC---CTTGAATTAAGATGAACAAACATTCAGACAGATCAGAT----- 2450
 Db 495 ValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
 QY 2451 -TTCCCTTCAGATCAAAACAAGAGTTGAGAAA---TTCTGGGATCTTGAAGTCT- 2505

Db 513 tPheProSerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGlu-SerL 533
 QY 2506 --CGGTGGACTGTTTCACAGAGGATGGTGTGTACC---CAGGCTACACACAAAGAAA 2560
 Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
 QY 2561 GGATAAAAT---AGTGGAAATTAAGAATTCATCAGCTATCAAAAT---CTTGTGACA 2614
 Db 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
 QY 2615 GTTCAT---CTTGTCAAAGGCAAGGAACTTCAAAAAGACACACTGTGAACA-CGTACAGGA 2670
 Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
 QY 2671 AAATGGCAACAT---GAAAGAGATTTTGTGCTGTAAGAAAGAAC---TGTGAGACAAA 2724
 Db 593 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 612
 QY 2725 AGAAATAAATCACACTAGAGAA---CCAAAAGTTAAATGGAAACAGAG---TCTGCAGT 2778
 Db 613 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnLeuLysCysSerVa 632
 QY 2779 GTAGATTCACTTAACCAA---GAAAAGAGAGAGAGAAATGCCGATATATTAAAGAAA 2835
 Db 632 lArgLeuThrLeuAsnGlnGluGluLysAsg-ArgAsnAlaAspIleLeuAsnGluL 652
 QY 2836 AA-----TTAGGAGAAATTAGAGATCGAGACGACATAGAAAGTTAGAA 2883
 Db 652 ysIleArgGluGluLeuGlyArgIle-GluGluGlnHisArg-----LysGluLeuGlu 669
 QY 2884 GTCAACACAACTTGAAG---GCTCTCAGATACAAG---ATAAGAAATTCAAGGTGTAGA 2937
 Db 670 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValG 689
 QY 2938 AAGTAATTCAATAGGTTCTCA---ACTCATGAATGAAA---ATTATCTTCAATGAA 2991
 Db 689 uSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluA 709
 QY 2992 ATTGCATGTTGAAAAAGGAATGCCA---TGCAAAACTGGAATAGCCACTA---AACA 3045
 Db 709 snCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis 728
 QY 3046 CCAATCAGGAAAGAAATAATC-TTTGAGGACTTAAGATTTTAAAGA---AAAGAT 3101
 Db 729 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsn 748
 QY 3102 GCTGAACCTTAG-ATGACCCCTAAACTGAAAGGAATCATTAATAAG---GGATCTCAA 3157
 Db 749 AlaGluLeuGlnMetThrLeuLysLysLysGluGluSerLeuThrLysArgAlaSerGln 768
 QY 3158 TATGTGGGAGCT---AAAGTTCTGATAGTACGAAACAATGCTCTTAATTAAGGA 3214
 Db 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 788
 QY 3215 AAAACAGACAAAGAAA-----CTAGAGCGCAATTCGAATA-CACCATCTGACTGCG 3267
 Db 788 ysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisPro----- 805
 QY 3268 TTCGCTGTACAAGC-----CATGATCAATTTGTGACATCAAGAAAAAGTAAGA- 3314
 Db 806 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 825
 QY 3315 --ACCTGCTCCACATTTAG---GAGATGCTGTTTGGCAAGAAATGAATGATGTGATGG- 3368
 Db 825 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 845
 QY 3369 --AGAGTACGATTAATAACAATGAGTGTCTCCATCACCACCTTTC---TGAGCTCAAGGAAA 3423
 Db 845 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
 QY 3424 TCAAAACCTAAAATTACT---CAATTAGCAGAGATGCTTAAGAGA---ATAACATT 3477
 Db 865 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 884

QY 3478 GGTTCAGAACATCACAAAGAGA---CAACGTGAAAAACAGTGTCAAT---GAAGGAGCG 3531
 Db 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904
 QY 3532 AACACATGTATCAAAAC-----GACAAGATATATGAAACAAACACCTGAACAGCAGATC 3585
 Db 904 luHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerL 924
 QY 3586 TCTAGTCAGAAATATTTCACACTACACACARAAATTTGGCTTTCACAGCAATT---AGT 3641
 Db 924 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuVal 943
 QY 3642 CATGCACATAGAAAGCTGCACAAACAAAGCAGACATAACAAT-----GATATT 3686
 Db 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
 QY 3687 CATTTCTTG---AGAGGAAATGCAACACATCTCTTAAAGAGAAAAGAGAGATATT--- 3741
 Db 960 HisPheLeuGluArgLysMetGlnHisLeuLeuLysGluLysAsnGlu-GluIlePh 979
 QY 3742 -AATTCAATACCATTTTAAAAACCGTATATT---CAATATCAAAAAGAGAAA 3789
 Db 979 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGlnLysGluLys 996
 RESULT 11
 AAU33357
 ID AAU33357 standard; protein; 1095 AA.
 AC AAU33357;
 DT 18-DEC-2001 (first entry)
 XX
 DE Human breast cancer protein B726P fusion protein #1.
 KW Human; ss; breast cancer protein; tumour; cancer; cytostatic;
 KW gene therapy; immunogen.
 OS Homo sapiens.
 XX
 PN WO200179286-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US012164.
 XX
 PR 17-APR-2000; 2000US-00551621.
 PR 08-JUN-2000; 2000US-00590751.
 PR 22-JUN-2000; 2000US-00604287.
 PR 20-JUL-2000; 2000US-00620405.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
 XX
 DR WPI; 2001-611721/70.
 DR N-PSDB; AAS47421.
 XX
 PT Breast Tumor Proteins and nucleic acids useful for the prevention,
 PT diagnosis and treatment of breast cancer.
 XX
 PS Claim 22; Page 292-295; 297pp; English.
 XX
 CC The invention relates to isolated breast tumour proteins and nucleic
 CC acids that encode them, including immunogenic fragments of the proteins.
 CC Also included are expression vectors expressing the proteins, transformed
 CC cells and antibodies raised against the proteins or an antigen presenting
 CC cell expressing the protein. The proteins and nucleic acids may be used
 CC in the prevention, diagnosis and treatment of diseases associated with
 CC inappropriate breast tumour protein expression, i.e. breast tumours and
 CC breast cancer e.g by gene therapy. The nucleic acids and their
 CC complements may also be used as DNA probes in diagnostic assays to detect
 CC and quantitate the presence of similar nucleic acids in samples, and

Db 686 LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysLysLeuSerGluAlaLys 705
 QY 2725 AGAAATAAATCACAGTAGAGAA---CCAAAGATTAAATGGCAACAGAG---TCTCAGT 2778
 Db 706 GluLeuLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnLeuLysCysSerVa 725
 QY 2779 GTAGATTGACTTAACCAA---GAAAGAGAAGAGAAATGCCGATATATTAAAGAAA 2835
 Db 725 lArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGluL 745
 QY 2836 AA-----TTAGGAGATTAGAAAGATCGAGACGACATAGAAAGATTAGAA 2883
 Db 745 yslleArgGluLeuLysGlyArgIle-GluGluGlnHisArg-----LysGluLeuGlu 762
 QY 2884 GTGAACACACAACTTGAAG---GCTCTCAGATCAAG---ATAAGAAATTGAAGTGTAGA 2937
 Db 763 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValG 782
 QY 2938 AAGTAATTGCAATAGTTTCTCA---ACTCATGAATGAAA---ATTATCTTACATGAA 2991
 Db 782 uSerAsnLeuAsnGln-VaSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGlu 802
 QY 2992 ATTGCATGTTCAAAAGAAATTGCCA---TGCACAACTGGAATAGCACACACTA---AACA 3045
 Db 802 snCysMetLeuLysLysGluLeuAlaMetLeuLysLeuGluIleAlaThrLeuLysHis 821
 QY 3046 CCAATCCAGGAAAGAAATAATC-TTTGAGGACTTAAAGATTTTAAAGA---AAAGAT 3101
 Db 822 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysLysLysLysLysAsn 841
 QY 3102 GCTGAACCTTAG-ATGACCTTAAACTGAAAGCAATCACTACTATAAG---GGATCTCAA 3157
 Db 842 AlaGluLeuGlnMetThrLeuLysLysLysLysLysLysLysLysLysLysLysLys 861
 QY 3158 TATGTGGCGAGCT---AAAGTCTGTAGCTGAGAAACAATGCTCTCTTAAATGAGAA 3214
 Db 862 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeu 881
 QY 3215 AAAACAGACAAAGAAA-----CTAGAGCGAAGAAATGAAATCA-CACCATCTGCTGCTG 3267
 Db 881 ysGluLysGlnAspLysGluLeuLeuGluAlaGluIleGluSerHisHisPro----- 898
 QY 3268 TTCGCTGTACAAGC-----CATCATCAATTGTGCATCAAGAAAGAAAGAAAGAA 3314
 Db 899 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGln 918
 QY 3315 --ACCTGCTTCCATGTAG--GAGATGCTGTTTCCAAAGAAATGAATGTGATGTG- 3368
 Db 918 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValaspVal 938
 QY 3369 --AGAGTACGATTAACAAATAGTGTCTCCATCCACCTTTC---TGAGCTCAAGAGAA 3423
 Db 938 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 957
 QY 3424 TCCAAACCTTAAATATCT---CAATPAGCAGGAGATGCTTAAGAGA---AATPACATT 3477
 Db 958 SerLysSerLysLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 977
 QY 3478 GGTTCCTCAGAACATCACAAAGAGA---CAACGTGAAACACAGTGTCAAT---GAAGAACGC 3531
 Db 977 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 997
 QY 3532 AACACATGATCAAAAC-----GACAAGATATGAAACAAACACCTGAACAGCAGATC 3585
 Db 997 luHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisHisThrGluGlnGlnSerL 1017
 QY 3586 TCTAGTCAAGAAATTTTCACTACAGCAAAATTTGGCTTCACAGCAATT-----AGT 3641
 Db 1017 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuVal 1036
 QY 3642 CATGCATAGAAAGCTGACACAAAGCAGATACAAAT-----GATATT 3686
 Db 1037 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 1052

QY 3687 CATTTCTTG---AGAGAAATGCACACATCTCTCTAAAGAGAAAGAGAGATATT-- 3741
 Db 1053 HisPheLeuGluArgLysMetGlnHisLeuLeuLysGluLysAsnGlu-GluIlePh 1072
 QY 3742 -AATTACAATACCATTTTAAACACCGTATATT---CAATATCAAAAGAGAAA 3789
 Db 1072 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGluLysGluLys 1089

RESULT 12

ABG78924
 ID ABG78924 standard; protein; 1095 AA.

XX ABG78924;

XX 15-NOV-2002 (first entry)

XX Human breast tumour polypeptide #15.

XX Human; breast tumour protein; breast cancer; cytostatic; vaccine.

XX Homo sapiens.

XX US2002085998-A1.

XX 04-JUL-2002.

XX 13-APR-2001; 2001US-00834759.

XX 28-DEC-1998; 98US-00222575.

XX 02-APR-1999; 99US-00285480.

XX 23-JUN-1999; 99US-00339338.

XX 02-SEP-1999; 99US-00389681.

XX 03-NOV-1999; 99US-00433826.

XX 17-APR-2000; 2000US-00551621.

XX 08-JUN-2000; 2000US-00590751.

XX 22-JUN-2000; 2000US-00604287.

XX 20-JUN-2000; 2000US-00620405.

XX (CORI-) CORIXA CORP.

XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;

XX Henderson RA;

XX WPI; 2002-635657/68.

XX N-PSDB; ABS64022.

XX Novel breast cancer polynucleotides and polypeptides encoded by the polynucleotides, useful for detecting the presence of breast cancer in a patient, and in pharmaceutical compositions, for treating breast cancer.

XX Disclosure; Page 223-225; 247pp; English.

XX The invention relates to an isolated breast tumour polynucleotide and the polypeptide it encodes. The polynucleotide and polypeptide are useful for detecting the presence of breast cancer in a patient, and in pharmaceutical compositions for treating breast cancer. The sequences are useful for stimulating an immune response in a patient and can therefore be used in production of vaccines. The sequences are also useful for detecting the presence of a cancer in a patient, by obtaining a biological sample from the patient, contacting the biological sample with a composition of the invention and detecting the amount of polynucleotide that hybridizes to the sample. This sequence represents a human breast tumour polypeptide of the invention

XX Sequence 1095 AA;

Alignment Scores:
 Pred. No.: 5,81e-95 Length: 1095
 Score: 914.50 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.91% Indels: 145

DB: 5 Gaps: 83
 US-09-602-362E-22 (1-4115) x ABC78924 (1-1095)

QY 1031 ATGAGTCCCGAAGAAACACTGAGAA---ATTACGTGGCACA-AAAGGAGACCTAGG 1086
 DB MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 113
 QY 1087 AAGAC-GCATGGGAGAAAAGAA---CCCTGTAAAGATGGATCGCT-GCAAGAGTAACA 1141
 DB LysileAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 133
 QY 1142 TCTAATAAATAAAG---TTTGTAAAGAAATCTAAGATGTT-CCATGTCACAAAG 1197
 DB SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 153
 QY 1198 ---AATCATCTACAAGCAAGTCGCATGATCAGAGG---TCCCATCAGATCCAAAACAAG 1251
 DB GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 173
 QY 1252 GAAGATCAAGATATCTTT-GATTCCTGGGTCTCTTTG-----AAGTCTGCAAGATT 1304
 DB GluAspGluGluTrpSerCysAspSerArg---SerLeuPheGluSerSerAlaLysile 192
 QY 1305 CAAGTG---GTATACCTGAGCTCTATATAAAGTAATGAGATAAATAAGAG---TAG 1358
 DB GlnValCysileProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 212
 QY 1359 AAAGCCTCTAAAGCCATCTGCTTCAAGCCGCAATGMAAGCAAAA---CTCTTCCA 1415
 DB GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 232
 QY 1416 AATAAGCCTTTG-----AATGAAGATGAACAAACATTAACACAGATCCGTG-TTC 1465
 DB AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 250
 QY 1466 CCACCG---AATCCAAACAAAGCACTATAAGAAATCTTGAGATCTAGACTCTG--- 1520
 DB ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 270
 QY 1521 -GAGACTGTTTACAGAGG---ATTGTGTTTACCAAGGTTACACATCAACAAAAGATAGAT 1576
 DB sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 290
 QY 1577 AAAAA---ATGAAAA---TAGAAGTCCCTAATAAGTGGTCTCTGAGGCTACT 1630
 DB 290 spLysIleAsnGlyLysLeuGluLysProAsnLysAspGlyLeuLeuLysAlaThrC 310
 QY 1631 CGGAAT---GAAATTTCTATTCTCACTAAAGCCT---AGAATTGAACACATGCAAACTTC 1684
 DB 310 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 329
 QY 1685 AAGC-GAGCCTCCGGAAGC---CATCTGCTTCGAGCCTCCACTGAATC---AAAGT 1737
 DB LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
 QY 1738 CTGTCCCAATATAGCCTTGAATGAAATGAAAAACATGGAGGAGATGAGAA---CTC 1794
 DB 350 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluLeu 369
 QY 1795 CCATCAGATCCAAACAAAGCACTATAGAAAA---TTCTGGATCTGGATCTCTG--- 1849
 DB 370 ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 389
 QY 1850 -TAGACTGTTTACAGAGGATTG-TGTTTACCAAGG---CTCGCTCAAAAAGATAGA 1904
 DB 389 sGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluLeu 409
 QY 1905 TAAATATGAAATAGAGGTT---CCCTGTTAAAGTGGTCTCTGAGGCTAACT--- 1959
 DB 409 pLysIleAsnGlyLysLeuGluGluSerProValLysAspGlyLeuLeuLysAlaAsnCy 429
 QY 1960 -GGGAATGAAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGATGATGCAAACTTCAA 2015

DB 429 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 449
 QY 2016 AG---CAGGCTCCGAGAGCCATCGCTTCCAGGCTCCATTGAAA---TCAAAGTCT 2069
 DB 449 YsAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 469
 QY 2070 GTCCAAATAAAC---CTTGAATTAAGAATGAACAAACATTCAGGCAGATGAGAA---CT 2123
 DB 469 alProAsn-LysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 488
 QY 2124 CCATCA---AATCCAAACAAAGGACTATGAGAAAGTCTCTGGATTCCTGGAGTCTCTG- 2179
 DB 488 uproSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeu 508
 QY 2180 --TAGACTGTTTCCAGAGGATGGTG---TTTACCAGGCTACACATCAAAAAGATAGA 2234
 DB 508 ysGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleA 528
 QY 2235 TAAAA---TAATGAAATAGAGACTCTCT-GATAATGAGTCTTCTGAAGCTCCCT- 2289
 DB 528 spLysIleAsnGlyLysLeuGluGluSerProAsnAspGlyPheLeuLysAlaProC 548
 QY 2290 --GAGATGAAAGTCTTATCCACTAAAGCCT---TGAATTGATGCGCATGCAAACTTCA 2344
 DB 548 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 568
 QY 2345 AAGCAGCCTCCG---AGAGCCATCTGCTTCCAGGCTGCTGAAAT---CAAAGTCT 2398
 DB 568 ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 587
 QY 2399 GTTCCAAATAAAC---CTTGAATTAAGAATGAACAAACATTCAGACACATCAGAT----- 2450
 DB 588 ValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArgAlaAspGlnMe 606
 QY 2451 -TTCCCTTCAGATCAAAACAAAGAGTGTAGAAA---TTCTGGATTCCTGAGTCT- 2505
 DB 606 tPheProSerGluSerLysGlnLys---ValGluGluAsnSerTrpAspSerGlu-SerL 626
 QY 2506 --CCGTGACTGTTTACAGAGGATGGTGTGTAC---CAGGCTACACAAAAAGAAA 2560
 DB 626 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 645
 QY 2561 GGATAAAT---AGTGGAAATTAGAGATTCACATAGCCTATCAAAAT---CTTGGTACA 2614
 DB 646 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 665
 QY 2615 GTTCAT---CTTGTGAAAGGCAAGGCACTTCAAAAAGCACTGTGAACA-CGTACAGGA 2670
 DB 666 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 685
 QY 2671 AATGGNAACAAT---GAAAAGAGTTCCTGTACTGAAAAGAAC---TGTACAGACAA 2724
 DB 686 LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysSerGluAlaLys 705
 QY 2725 AGAAATAAATACAGATAGAGAA---CCAAAAGTTAAATGGGAACAGAG---TCTGCAGT 2778
 DB 706 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerVa 725
 QY 2779 GTAGATTGACTTAAACAA---GAAAAGAGAGAGAGAAATCCGATATATTAAGAAAA 2835
 DB 725 laArgLeuThrLeuAsnGlnGluGlnLysArg-ArgAsnAlaAspIleLeuAsnGluL 745
 QY 2836 AA-----TTAGGAAGAATTAGAAGAAATCGAGACGATAGAGAGATGAGAA 2883
 DB 745 ysIleArgGluGluLeuGlyArgIle-GluGluGlnHisArg-----LysGluLeuGlu 762
 QY 2884 GTGAACACAACTTGAAG---GCTCTCAGATACAG---ATAAGAAATGAGGTCTAGA 2937
 DB 763 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGluAspIleGluLysSerValG 782
 QY 2938 AAGTAATTTCAATAGGTTTCTCA---ACTCACTGAATGAAA---ATTATCTTACATGAA 2991

Db 782 userAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluA 802
 QY 2992 ATTGCATCTTCAAAAGGAATGCGCA---TGCAAAACCTGGAATAGCCACACTA---AACA 3045
 Db 802 snCysMetLeuLysGluLeuAlaMetLeuLysGluLeuAlaThrLeuLysHis 821
 QY 3046 CCAATCCAGGAAAGGAAATAATATC-TTTGAGGACTTAAGATTTTAAAGA---AAAGAT 3101
 Db 822 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsn 841
 QY 3102 GCTGAACCTAG-ATGACCCCTAAATCTGAAAGGAAATCATTTACTAAAG---GGATCTCAA 3157
 Db 842 AlaGluLeuGlnMetThrLeuLysLeuLysGluLysSerLeuThrLysArgAlaSerGln 861
 QY 3158 TATCTGGGCAGCT---AAAGTTCTGATAGCTGAGAAACAATGCTCATTTCTAAATTCAGGA 3214
 Db 862 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 881
 QY 3215 AAAACACACAAAGAAA---CTAGAGCAGAAATTGATA-CACCATCTCTGACTGGC 3267
 Db 881 ysGluLysGlnAspLysGluLeuGluAlaGluIleGluSerHisHisPro----- 898
 QY 3268 TTGCTCTACAGC-----CATGATCAATTGTGACATCAAGAAAGTAAGA- 3314
 Db 899 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 918
 QY 3315 --ACCTGCTTCCACATTGAG---GAGATGCTGTTTCCAAAGAAATGAATGTGATGTG- 3368
 Db 918 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 938
 QY 3369 --AGAGTACGATATTAACATGATGCTCCATCACCACATTC---TGACCTCAAGGAAA 3423
 Db 938 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 957
 QY 3424 TCCAAACCTTAAATATCT---CAATTAGCAGGAGATGCTCTAAGAGA---AATACATT 3477
 Db 958 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrIe 977
 QY 3478 GGTTCAGACATCACAAAGAG---CAAGCTGAAACACAGTGTCAAT---GAAGGAAGC 3531
 Db 977 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 997
 QY 3532 AACACATGATCAAAAC-----GACAAGATATGAAACAAACACCTGACACAGCAGATC 3585
 Db 997 luHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGlnGlnGlnGluSerL 1017
 QY 3586 TCTAGTCAGAAATATTTCACATACAGCAAAATTTGGCTTCACAGCAATT---AGT 3641
 Db 1017 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 1036
 QY 3642 CATGCACATAGAAAGCTGACAAACAAAGCAGATAACAAT-----CATATT 3686
 Db 1037 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 1052
 QY 3687 CATTTCTTG---AGAGAAATCAACACATCTCTCTAAAGAGAAAGAGGAGATATT-- 3741
 Db 1053 HisPheLeuGluArgLysMetGlnHisLeuLeuLysGluLysAsnGlu-GluIlePh 1072
 QY 3742 -AATTACATACCATTTTAAARCCCTATATT---CAATATGAAAGAGAA 3789
 Db 1072 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGlnLysGluLys 1089
 RESULT 13
 ABJ37747
 ID ABJ37747 standard; protein; 1095 AA.
 XX
 AC ABJ37747;
 XX
 DT 15-MAY-2003 (first entry)
 XX
 DE Human tumour-related protein - SEQ ID No 493.
 XX
 KW Human; vaccine; gene therapy; T cell stimulation; T cell expansion;

KW tumour; breast cancer; cancer; immune response stimulation.
 XX Homo sapiens.
 OS WO200283956-A1.
 PN 24-OCT-2002.
 XX 15-APR-2002; 2002WO-US012378.
 PF 13-APR-2001; 2001US-00834759.
 PR 07-DEC-2001; 2001US-00007805.
 PR 13-FEB-2002; 2002US-00076622.
 XX (CORI-) CORIXA CORP.
 PA Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 XX Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
 PI Vedwick TS, McNeill PD, Durham M;
 XX WPI; 2003-103376/09.
 XX New polypeptide and polynucleotide useful for stimulating and/or
 PT expanding T cells specific for a tumor protein and treating breast
 PT cancer.
 PS Disclosure; Page 310-312; 375pp; English.
 CC The invention comprises a method of stimulating and/or expanding T cells
 CC specific for a tumor protein. The invention further comprises human
 CC nucleic acids and proteins that are associated with tumors (e.g. breast
 CC cancer). The method and sequences of the invention are useful for
 CC stimulating and/or expanding T cells specific for a tumor protein,
 CC detecting the presence of cancer, stimulating an immune response in a
 CC patient and treating breast cancer. The present amino acid sequence
 CC represents a human tumour-related protein
 XX Sequence 1095 AA;
 SQ
 Alignment Scores:
 Pred. No.: 5,81e-95 Length: 1095
 Score: 914.50 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.91% Indels: 145
 DB: 6 Gaps: 83
 US-09-602-362E-22 (1-4115) x ABJ37747 (1-1095)
 QY 1031 ATGAGTCCCGAAGAAACACTGAGAA---ATTACCTGGGCGA-AAAGAGACCTAGG 1086
 Db 94 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 113
 QY 1087 AAGAC-GCATGGGAGAAAGAA---CCCTGTAAAGATGGATGCGT-GCAAGAGTAACA 1141
 Db 114 LysIleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThr 133
 QY 1142 TCTAATAAATAAG---TTTGAAGAAAGAAATCTAAGATGTT-GCATGTCCACAAAG 1197
 Db 134 SerAsnLysThrLysValLeuGluLysGlyA-gserLysMetIleAlaCysProThrLys 153
 QY 1198 ---AATCATCTCAAAAGCAAGTCCATGATCAGAG---TCCCATCAGATCCAAACAGG 1251
 Db 154 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 173
 QY 1252 GAAGATGAAGAATATCTTTT-GATTCTCGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
 Db 174 GluAspGluGluTyr-SerCysAspSerArg---SerLeuPheGluSer-SerAlaLysIle 192
 QY 1305 CAAGTG---GTATACCTGAGTCTATATATAAAGAACTAAGATGAATAAAGAG---TAG 1358
 Db 193 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 212

1359 AAAGCCTCTAAAGCCATCTGCTTCAAGCGCGCATTTGAAGCAAAA---CTCTTTCCA 1415
Db
213 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 232
Qy
1416 AATAAGCCTTTG-----AATGAGATGATGAACAATGTAAGCAGATCCGTG-TTC 1465
Db
233 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 250
Qy
1466 CCACCG---AATCCAAACAAAGGACTATAAGAAATTTCTGGATTCTAGAGTCTCTG-- 1520
Db
251 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 270
Qy
1521 -GAGACTGTTTACAGAGG---ATTGTGTTTACCAAGGCTACACATCAAAAAGAATAGAT 1576
Db
270 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 290
Qy
1577 AAAAA---AATGAAAA---TAGAGAGTCCCTAATAAGTGTCTTCTGAAGGCTACT 1630
Db
290 sPlysIleAsnGlnLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 310
Qy
1631 CGGAAT---GAAATTTCTATTCCACTAAAGCCT---AGAAATTAAGACATGCAAACTTTC 1684
Db
310 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 329
Qy
1685 AAGC-GAGCCTCCGGAAGC---CATCTGCTTCGAGCCTCCACTCAATC---AAGT 1737
Db
330 LysAlaGluProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
Qy
1738 CTGTCCCAATAAGCCTTGAATGAAATGAAAAATGAAATGAGGAGGAGATGAGAA---CTC 1794
Db
350 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 369
Qy
1795 CCATCAGATCCAAACAAAGGACTATAAGAAA---TTCTGGATCTGAGTCTCTG-- 1849
Db
370 ProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSer-GluSerLeuCy 389
Qy
1850 -TAGACTGTTTACAGAGGATG-TGTTTACCAAG---CTCGCTCAAAAAGAAATAGA 1904
Db
389 sGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleAs 409
Qy
1905 TAAATAATGGAATATTAGAAGGT---CCCTGTTAAAGTGTCTTCTGAGGCTACT-- 1959
Db
409 pLysIleAsnGlnLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 429
Qy
1960 -GGATGAAAGTCTTATCCACTAAGCCTAGA---ATTGATGATGATGCAACTTCAA 2015
Db
429 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPhe 449
Qy
2016 AG---CAGGCTCCCGAGAGCCATCTGCTTCCAGCTCCATTGAAA---TCAAAAGTCT 2069
Db
449 ysAlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 469
Qy
2070 GTCCAAATAAAC---CTTGAATTAGAATGAACAACATGAGGAGATGAGAA---CT 2123
Db
469 alProAsn-LysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 488
Qy
2124 CCATCA---AATCCAAACAAAGGACTATGAGAAAGTCTTGGCAATCTGAGTCTCTG- 2179
Db
488 uproSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeu 508
Qy
2180 --TAGACTGTTTCCAGAGGATGGT---TTTACCAGGCTACACATCAAAAAGAATAGA 2234
Db
508 ysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleA 528
Qy
2235 TAAA---TAATGGAATAGAGAGTCTCT-GATATGAGTGTCTTCTGAAGCTCCCT- 2289
Db
528 sPlysIleAsnGlnLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProC 548
Qy
2290 --GAGAATGAAAGTCTTATCCACTAAAGCCT---TGAATTGATGCGATGCAAACTTCA 2344
Db
548 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPhe 568
Qy
2345 AAGCAGCCTCCCG---AGAGCCATCTGCTTCGAGCCTGCAATGAAT---GAAAAGTCT 2398

568 ysAlaGluProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 587
Qy
2399 GTTCCAAATAAAC---CTTGGAAATTAAGATGAACAACATTGAGACAGATCAGAT----- 2450
Db
588 ValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArgAlaAspGlnMe 606
Qy
2451 -TTCCCTTCAGATCAAAAACAAGAGGTTGAGAAA---TTCTGGATTTCTGAGTCT- 2505
Db
606 tPheProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGlu-SerL 626
Qy
2506 --CGCTGACTGTTTACAGAGGATGCTGTGTACC---CAGGCTACACACAAAAAGAAA 2560
Db
626 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 645
Qy
2561 GGATPAAAT---AGTGAATAATAGAAAGATCTACAGCTTATCAAAAT---CTTGTGACA 2614
Db
646 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 665
Qy
2615 GTTCAT---CTTGTGAAGGCAAGGAACTTCAAAAAGACACTGTGAACA-CGTACAGGA 2670
Db
666 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 685
Qy
2671 AAATGGAAACAT---GAAAAGAGTGTGTTGTACTGTAAGAAAGAAC---TGTCAAGACAA 2724
Db
686 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 705
Qy
2725 AGAATAATACACAGTAGAGNA---CCAAAGTTAAATGGGAACAAGAG---TCTGCAGT 2778
Db
706 GluIleLysSerGln-LeuGluAsnGlnLysValLysValLysIleGlnGluLeuCysSerVa 725
Qy
2779 GTAGATGACTTAAACCAA---GAAAAGAGAGAGAAATGCCGATATATTAAGAAAA 2835
Db
725 lArgLeuThrLeuAsnGlnGluGluLysArg-ArgAsnAlaAspIleLeuAsnGlu 745
Qy
2836 AA-----TTAGGAAGATTAGAAATTCGAGAGCAGCAGCATAGAAGAGTTAGAA 2883
Db
745 ysIleArgGluGluLeuGlyArgIle-GluGluGlnHisArg-----LysGluLeuGlu 762
Qy
2884 CTGAACACACTTGAAG---GCTCTCAGATACAAG---ATAAGATTGAAGGTGTAGA 2937
Db
763 ValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValG 782
Qy
2938 AAGTAATTTGAATAGTGTCTCA---ACTCATGAAATGAAA---ATTATTTCTTACATGAA 2991
Db
782 uSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluA 802
Qy
2992 ATTGATGTTTGAAGAGGAAATGGCA---TCCAAACTGGAATAGCCACACTA---AAAC 3045
Db
802 snCysMetLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis- 821
Qy
3046 CCAATCCAGGAAAAGGAAAATAAATC-TTTGAGGACTTAAGATTTTAAAGA---AAAGAT 3101
Db
822 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsn 841
Qy
3102 GCTGAACCTTAG-ATGACCTTAAACTGAAAAGGATCATTTACTAAAG---GGATCTCAA 3157
Db
842 AlaGluLeuGlnMetThrLeuLysLeuLysGluLysLeuThrLysArgAlaSerGln 861
Qy
3158 TATGTGGCAGCT---AAAGTTCTCATAGCTGAGAAACAATGCTCATTTCTAAATTGAGA 3214
Db
862 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 881
Qy
3215 AAAACAGACAAGAAAA-----CTAGAGGAGAAATTTGAATA-CACCATCTGACTGCGC 3267
Db
881 ysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisPro----- 898
Qy
3268 TTCGCTGTACAGC-----CATGATCAATTGTGTGACATCAAGAAAAAGTAAGA- 3314
Db
899 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 918
Qy
3315 --ACCTGTCTCCACATTGAG---GAGATGCTGTTTTCAAAGAAAATGAATGTTGATGTG- 3368

Db 918 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 938
 QY 3369 --AGAGTACGATATTAACAATGAGTGTCCATCACCACCTTTC---TGAGCTCAAGGAAA 3423
 Db 938 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 957
 QY 3424 TCCAAAACCTAAAATATATCT---CAATTAGCAGGAGATGCTTAAGAGA---AATACATT 3477
 Db 958 SerLysSerLeuLysIleAsnLeuAsnTyrAlaGlyAsp-AlaLeuArgGluAsnThrLe 977
 QY 3478 GGTTCAGACATCACAAAAGAGA---CAACGTGAAAACAGTGTCAAT---GAAGGAAGCG 3531
 Db 977 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 997
 QY 3532 AACCATGTATCAAAAC-----GACAAGATAATGGAACAAACACCTGAAACAGCAGTGC 3585
 Db 997 luHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerL 1017
 QY 3586 TCTAGTCAGAAATATTTCACACTACAGCAAAATTTGCGCTTCACAGCAATT---AGT 3641
 Db 1017 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 1036
 QY 3642 CATGCACATAGAAAGCTGCACAAACAAACAGCAGATAACAAT-----GATATT 3686
 Db 1037 HisAlaHisLys-----LysAlaAsnLysSerLysIleThrIleAspIle 1052
 QY 3687 CATTCTCTG---AGAGAAATGCAACACATCTCTCTAAAGAGAAAAGAGGAGATATT--- 3741
 Db 1053 HisPheLeuGluArgLysMetGlnHisHisLeuLeuLysGlnLysAsnGlu-GluIlePh 1072
 QY 3742 -AATTACAATACCATTTTAAACACCGTATATT---CAATAGAAAAGAGAAA 3789
 Db 1072 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGluLysGluLys 1089
 RESULT 14
 ABJ37789
 ID ABJ37789 standard; protein; 1239 AA.
 XX AC ABJ37789;
 XX DT 15-MAY-2003 (first entry)
 XX DE Human tumour-related protein - SEQ ID No 577.
 XX KW Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
 XX KW tumour; breast cancer; cancer; immune response stimulation.
 XX OS Homo sapiens.
 XX PN W0200283956-A1.
 XX PD 24-OCT-2002.
 XX PF 15-APR-2002; 2002WO-US012378.
 XX PR 13-APR-2001; 2001US-00834759.
 XX PR 07-DEC-2001; 2001US-00007805.
 XX PR 13-FEB-2002; 2002US-00007622.
 XX PA (CORI-) CORIXA CORP.
 XX PI Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
 PI Vedvick TS, McNeill PD, Durham M;
 XX DR WPI; 2003-103376/09.
 XX PT New polypeptide and polynucleotide useful for stimulating and/or
 PT expanding T cells specific for a tumor protein and treating breast
 PT cancer.
 XX PS Example 13; Page 359-363; 375pp; English.

CC The invention comprises a method of stimulating and/or expanding T cells
 CC specific for a tumour protein. The invention further comprises human
 CC nucleic acids and proteins that are associated with tumours (e.g. breast
 CC cancer). The method and sequences of the invention are useful for
 CC stimulating and/or expanding T cells specific for a tumour protein,
 CC detecting the presence of cancer, stimulating an immune response in a
 CC patient and treating breast cancer. The present amino acid sequence
 CC represents a human tumour-related protein
 XX
 SQ Sequence 1239 AA;

Alignment Scores:
 Pred. No.: 3,76e-60 Length: 1239
 Score: 612.50 Matches: 458
 Percent Similarity: 45.75% Conservative: 166
 Best Local Similarity: 33.58% Mismatches: 385
 Query Match: 8.64% Indels: 363
 DB: 6 Gaps: 90

US-09-602-362E-22 (1-4115) x ABJ37789 (1-1239)

QY 380 CATGTGCGATATCGAAGTGCCTCAAC---CAAGCGTGCCTCACACACCTTTTACTACCAT-- 434
 Db 8 TyrGlyAlaValIleGluValGlnAsnLysAlaSerLeuThrProLeuLeuLeuAlaI 27
 QY 435 -AACGAAAGAGTGCAGCAATTGTGGAATTT---GCTGATAAAATGCAATGCGAATGC 490
 Db 27 eGlnLysArgSerLysGlnThrValGluPheLeuLeuThrLysAsnAlaAsnAla 47
 QY 491 GTTA---ATAAGTTAAATGCAACCCCTCATGCTCTGTATG---TCTGGATCATCAGAGA 544
 Db 47 aPheAsnGluSer-LysCysThrAlaLeuMetLeuAlaIleCysGluGlySerGluI 67
 QY 545 TAGTTGATGCTCTCTTAGCAAAA---TGTGACGCTTTTGTGCGAG---ATATAGTGGAGTA 598
 Db 67 leValGlyMetLeuLeuGlnGlnAsnValAspValPheAlaGluAspIle-HisGlyIle 86
 QY 599 ACTGCAGAA---CATATGCTGTTACTGTGATTTCTCACATTTCATAACAATATGAATA 655
 Db 87 ThrAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIleHisGlnGlnLeuLeuGlu 106
 QY 656 TAT---ACGAAATATTCAAAATCATCAATACCAATCAGAAAGAACTCTGCGAGGAAC 712
 Db 107 HisIleArgLysLeuProLysAsn----- 114
 QY 713 CTGATGAGCTGCACCCCTGGCGGAAAGACACCTGACACCGTGAAAGCTGTGGGAAAACA 772
 Db 115 -----ProGlnAsnThrAsnProGluGlyThrSerThrGly-----Thr 127
 QY 773 CCTGATGAGCTGCACCCCT---GGTGGAAAGACACCTGACAGGCTGAAAGCT---GGTGG 826
 Db 128 ProAspGluAlaAla-ProLeuAlaGluArgThrProAspThrAlaGluSerLeuLeuGlu 147
 QY 827 AAAAAGACCTGATGAG---CCTGCATCTTGTGGAGGAGACATCTGACAAATTCAT---C 880
 Db 147 uLysThrProAspGluAlaAla-ArgLeuValGluGlyThrSerAlaLysIleGlnCysL 167
 QY 881 TTGAGAAAGGACATCTGGAAGTTTCGAACCTCAGC---AGAAAAACACCTAGCAAAA--- 934
 Db 167 euGlyLysAlaThrSerGlyLysPheGluGln-SerThrGluGluThrProArgLysIle 186
 QY 935 TTACGATCTGCAAAAACATCTGAGAAATTTACGGG-CCAGCAAGAGACCTAGAGAA 993
 Db 187 LeuArgProThrLysGluThrSerGluLysPheSerTrpProAlaLysGlu----- 203
 QY 994 GATCGCAGGAGAAAAGAACACACCTAGGGAATATGATGCCGCAAGAAACACACTG 1053
 Db 204 -----ArgSerArgLysIleThr-----TrpGluGluLysGluThrSerVal 217
 QY 1054 AGAAATACGTGGCGAGAAAAGAACACCTAGGAACACGATCGGAGAAAAGAACCCCT 1113
 Db 218 LysThrGluCysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThr 237

1114 GTAAAGATGGATCGCTGCAAGAGTAACATCAATAAATAA----- 1155
Db |||||
238 SerAsnMetIleAlaCysProThrLysGluThrLysAlaSerThrAsnValAsp 257
QY |||||
1155 ----- 1155
258 ValSerSerValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGln 277
Db |||||
1156 ----- 1156
278 CysThrLysValGluAspPheAsnLeuAlaThrLysIleLeuSerLysSerAlaAla 297
Db |||||
1174 TAAGATGTTGATGT---CCACAAAGAATCAT-----CTACAAGC 1212
QY |||||
298 GlnAsnTyrThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHis 317
Db |||||
1213 AAGTCCCATGATCAGAGG---TCCCATCATGATCCAAACAAAGGAGATGAAGATATTCT 1269
QY |||||
318 LysIleGluAspGlnMetPheProSerGluSerLysArgGluGluAspGluGluTyrSer 337
Db |||||
1270 TTGATTCGGTCTCTTTGAAGTTCG---AAGATTCAGTG---GTATACCTGAGT 1322
QY |||||
338 TrpAspSerGlySerLeuPheGluSerAlaLysThrGlnValCysIleProGluSer 357
Db |||||
1323 CTATATATAAAGTAATGAGATAATAAGAG---TAGAAGCCCTCTAAAGCCATCT 1379
QY |||||
358 MetTyrGlnLysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSer 377
Db |||||
1380 GCCTTCAAGCCGCCATTGAAA---GCAAACTCTTTCCAAATPAGCCT---TTGAATGAA 1433
QY |||||
378 AlaPheLysProAlaValGluMetGlnLysThrValProAsnLysAlaPheGlu-LeuLys 397
Db |||||
1434 GAATGAAACAACTTGAACAGATCCGTGTT---CCCACGATCCAAACAAAGG---AC 1487
QY |||||
397 sAsnGluGlnThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAs 417
Db |||||
1488 TATAAGAAAATCTCTGGGATCTAGAGTCT---CTGGAGACTGTTTACAGAGAGGATTGT 1544
QY |||||
417 pGlu-GluAsnSerTrpAspSer-GluSerProCysGluThrValSerGlnLysAspVal 436
Db |||||
1545 TTT---ACCAAGCTCACATCAAAAAGAA---TAGATAAAAATGGAATAATAGAGAG 1598
QY |||||
437 TyrLeuProLysAlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGlu 456
Db |||||
1599 TCCTAATAAAGTGTCTCTGAAGCTACCTCGG-----AATCAAAATTTCT 1645
QY |||||
457 SerProValLys-----AspGlyLeuLeuLysProThrCysGlyArgLysValSer 473
Db |||||
1646 ATTCCCTAAAGCCT---AGAATTGAAGCATGCAAACTTTCAAAGCG----- 1690
QY |||||
474 LeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaGluSerProAsp 493
Db |||||
1691 -----AGCCTC----- 1696
QY |||||
494 LysAspGlyLeuLeuLysProThrCysGlyArgLysValSerLeuProAsnLysAlaLeu 513
Db |||||
1697 -----CGGAAGCCATCTGCTCGAGCCTCCA-----CTGAATCA 1732
QY |||||
514 GluLeuLysAspArgGluThrLeuLysAlaGluSerProAspAsnAspGly-LeuLeuLys 533
Db |||||
1733 AAAGTCTGT-----CCCAATAGCCTTGAATGAAATGAAATGAAAC 1774
QY |||||
534 ProThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGlu-----LeuLysAsp 551
Db |||||
1775 ATGAGGCAGNATGAA-----CTCCCATCAGAATCCAAACAAAGGACTATAAGA 1824
QY |||||
552 ArgGluThrPheLysAlaAlaGlnMetPheProSerGluSerLysGlnLysAspGlu 571
Db |||||
1825 AAA---TTCTGGGACTGAGTCTCTGTAGACTGTTTCAAGAGATTG----- 1872
QY |||||
572 GluAsnSerTrpAsp-----PheGluSerPheLeuGluThrLeuLeuGlnAsn 587
Db |||||
1873 -----TGTTTACCAAGG---CTCGGCTCAAAAGATAGAT-----AAAATA 1911
QY |||||

588 AspValCysLeuProLysAlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeu 607
Db |||||
1912 ATGGAAATATAGAGGTCCTCTTAAAGTGGCTTTCTGAGGCTAACT---GGGAATGAA 1968
QY |||||
608 GluGluSer-----ProAspLysAspGlyLeuLeuLysProThrCysGlyMetLys 624
Db |||||
1969 AGTTCATTCCAACTAAAGCCTAGA---ATTGATGACATGCCAACTTCAAGCAGGCTCT 2025
QY |||||
625 IleSerLeuProAsnLysAla-LeuGluLeuLysAspArgGluThrPheLys-----Al 642
Db |||||
2026 CCAGAGACCATCTGCTTCGAGCCTCCATTGAAATCAAAAGTCTGCCAAATAAACCTTG 2085
QY |||||
642 aGluAspValSerSerValGluSerThrPheSerLeuPheGlyLysProThrThrGlu-- 661
Db |||||
2086 GAATTAAAGATGAACAAACATTGAGGCAGAT-----GAGAACCTCCA 2127
QY |||||
662 -AsnSerGlnSerThrLysValGluGluAspPheAsnLeuThrLysGluGlyAlaTh 681
Db |||||
2128 TCAAAATCCAAACAAAGGACTATGAGAAAGTCTCTGGATTTCTGGAG----- 2173
QY |||||
681 rLysThrValThrGlyGlnGlnGluArgAspIleGlyIleIleGluArgAlaProGlnAs 701
Db |||||
2173 ----- 2173
701 pGlnThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThrSe 721
Db |||||
2174 -----TCTCTGTAGACTGTTCCAGAAAGGATGTTTACCAGGCTAC 2217
QY |||||
721 rAspSerGluIleIleSerValSerAspThrGlnAsnTyrGluCysLeuProGluAlaTh 741
Db |||||
2218 ACATCAAAAAGATAGATAAAA-----TAATGGAAAATAGAGAGCTCTGTGATATGAG 2271
QY |||||
741 rTyrGlnLysGlu---IleLysThrThrAsnGlyLysIleGluGluSer----- 756
Db |||||
2272 GTTTCTGAAGGCTCCCTCGAGATGAAGTTCATTCCACTAAAGCCTTGAATTGATGC 2331
QY |||||
756 ----- 756
2332 ATGCAAACTTTCAAAAGCAGGCTCCCGAGAGCCATCTGCTTCGAGCCTGCATTGAAATGA 2391
QY |||||
757 -----ProGluLysProSerHisPheGluProAlaThrGluMet-----G 770
Db |||||
2392 AAAGTCTGTTCCAAATAAA-----CCTTGAATTAAGAAATGAGAACATTTGACAGAT 2445
QY |||||
770 nAsnSerValProAsnLysGlyLeuGluIleLysAsnLysGlnThr----- 785
Db |||||
2446 CAGATTTCCCTTCAGAAATCAAAACAAAGAGGTTGAGAAAATTTCTGGGATTTCTGAAGTCT 2505
QY |||||
786 -----LeuArgAlaAspSerThrThrLeuSerLysIleLeu-AspAlaLeuProS 802
Db |||||
2506 CCGTGGACTGTTTACAGAGAGGATGGTGTGTACCCAGGCTACACACAAAAGAGGATA 2565
QY |||||
802 erCysGlu----- 804
Db |||||
2566 AAATAGTGGAAAATTAGAAGATTCCTAGCCTATCAAAATCTTGTGTACAGTTTCATTTGT 2625
QY |||||
804 ----- 804
2626 GAAAGGCAAGGGAACCTTCAAAAAGACACCTGTGCAACAGCTAC---AGGAAATGGAACAT 2682
QY |||||
805 -ArgGlyArgGluLeuLysLysAspAsnCysGluGlnIleThrAlaLysMetGluGlnM 824
Db |||||
2683 GAAAAGAGTTTTGTGTACT---GAAAAGAGCTGTGAGACAAAAGAAATAATCA- 2738
QY |||||
824 etLysAsnLysPheCysValLeuGlnLysGluLeuSerGluAlaLysGluIleLysSerG 844
Db |||||
2739 --GTAGAAACCAAAAGTT-AAATCGGAAACAAAGAG---TCTGCAGTGTGAGATTGACTTAA 2792
QY |||||
844 InLeuGluAsnGlnLysAlaLysTrpGluGlnLeuCysSerValArgLeuProLeuA 864
Db |||||
2793 ACCAA---GAAAAGAGAGAGAGAAATCCCATATATTAAAGAAAATAATAGGAGAAATT 2849
QY |||||

1186	QY	TG	TGCCA	CAAAAG	---	AAT	CAT	CTAC	AAAGCAAG	TGCCAT	GCAT	CAGAGG	---	TCCCAT	CAG	12339												
68	Db	Cys	Pro	Thr	lys	Glus	er	Thr	Lys	Ala	asn	Asp	gln	Arg	Phe	Pro	Ser	Glu	87									
1240	QY	AT	CAAA	CAAG	GAAG	AT	GA	AT	AT	CTTT	---	GAT	TCG	GGT	CT	CTTTG	---	AAG	1292									
88	Db	Ser	Lys	gln	Glu	asp	Glu	gln	u	tyr	Ser	Cys	asp	Ser	arg	---	Ser	Leu	Phe	Glu	106							
1293	QY	TT	CTG	CAAG	AT	CAAG	TG	---	GT	AT	AC	TG	AG	TCT	AT	ATA	AAAA	AGT	AA	TG	AT	GA	TA	AA	T	1349		
107	Db	Ser	Ala	lys	ile	gln	Val	Cys	ile	Pro	Glu	Ser	ile	tyr	Gln	lys	Val	Met	Glu	ile	Asn	126						
1350	QY	AAG	NAG	---	TAG	AA	AG	CT	CT	TA	AA	AG	CA	T	CG	CT	TC	AA	AG	CG	CA	T	TA	GA	AG	CA	AAA	1406
127	Db	Arg	gln	Val	gln	Glu	Pro	Pro	lys	Lys	Pro	Ser	Ala	Phe	lys	Pro	Ala	ile	Glu	Met	Gln	146						
1407	QY	---	CT	CTT	TCC	AA	T	AA	GC	CTT	TG	---	---	AAT	GA	AA	AT	GA	CA	CA	AT	TG	AA	GC	CA	1454		
147	Db	Asn	Ser	Val	Pro	Asn	lys	Ala	Phe	Glu	Leu	Lys	Asn	Glu	gln	---	Thr	Leu	Arg	Gla	164							
1455	QY	GAT	CCG	TG	---	TT	CC	C	ACG	---	---	AAT	CC	AA	CA	AA	AGG	CA	T	AT	AG	AA	AT	TT	TT	TT	TT	1510
165	Db	Asp	Pro	Met	Phe	Pro	Glu	Ser	Lys	gln	Lys	Asp	Tyr	Glu	Glu	Asn	Ser	Trp	Asp	Ser	184							
1511	QY	AG	AGT	CT	CTG	---	GAG	ACT	CT	T	T	CA	CA	AGG	---	---	ATT	GT	CT	T	T	CA	CA					

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:47:52 ; Search time 34.9972 Seconds
(without alignments)
12140.450 Million cell updates/sec

Title: US-09-602-362E-22
Perfect score: 7086
Sequence: 1 ctagtctatacagcaagac.....gagtgactcacctgaaaa 4115

Scoring table:

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US09602362/runat_15072004_093627_22040/app_query.fasta_1.10325
-DB=Issued Patents:AA -OFMT=fastan -SUPFIX=rai -MINMATCH=0.1 -IOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362 @cgn_1_133 @runat_15072004_093627_22040 -NCFU=6 -TCPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -HEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents:AA:*
1: /cgn2_6/prodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	917	12.9	1002	4	US-09-620-405B-475
2	917	12.9	1002	4	US-09-604-287A-475
3	917	12.9	1002	4	US-09-834-759-475
4	914.5	12.9	1095	4	US-09-620-405B-493
5	914.5	12.9	1095	4	US-09-834-759-493
6	583.5	8.2	650	4	US-09-620-405B-469
7	583.5	8.2	650	4	US-09-433-826B-469
8	583.5	8.2	650	4	US-09-604-287A-469
9	583.5	8.2	650	4	US-09-834-759-469
10	582	8.2	743	4	US-09-620-405B-494
11	582	8.2	743	4	US-09-834-759-494
12	421.5	5.9	445	4	US-09-620-405B-473

13	421.5	5.9	445	4	US-09-433-826B-473
14	421.5	5.9	445	4	US-09-604-287A-473
15	421.5	5.9	445	4	US-09-834-759-473
16	416.5	5.9	432	4	US-09-389-681-181
17	416.5	5.9	432	4	US-09-620-405B-181
18	416.5	5.9	432	4	US-09-339-338-181
19	416.5	5.9	432	4	US-09-433-826B-181
20	416.5	5.9	432	4	US-09-604-287A-181
21	416.5	5.9	432	4	US-09-285-480-181
22	416.5	5.9	432	4	US-09-834-759-181
23	408.5	5.8	466	4	US-09-620-405B-472
24	408.5	5.8	466	4	US-09-433-826B-472
25	408.5	5.8	466	4	US-09-604-287A-472
26	408.5	5.8	466	4	US-09-834-759-472
27	285	4.0	410	4	US-09-620-405B-495
28	285	4.0	410	4	US-09-834-759-495
29	279	3.9	317	4	US-09-389-681-176
30	279	3.9	317	4	US-09-620-405B-176
31	279	3.9	317	4	US-09-339-338-176
32	279	3.9	317	4	US-09-433-826B-176
33	279	3.9	317	4	US-09-604-287A-176
34	279	3.9	317	4	US-09-285-480-176
35	279	3.9	317	4	US-09-834-759-176
36	279	3.9	323	4	US-09-620-405B-483
37	279	3.9	323	4	US-09-604-287A-483
38	279	3.9	323	4	US-09-834-759-483
39	208.5	2.9	228	4	US-09-620-405B-470
40	208.5	2.9	228	4	US-09-433-826B-470
41	208.5	2.9	228	4	US-09-604-287A-470
42	208.5	2.9	228	4	US-09-834-759-470
43	203	2.9	179	4	US-09-389-681-179
44	203	2.9	179	4	US-09-620-405B-179
45	203	2.9	179	4	US-09-339-338-179

ALIGNMENTS

RESULT 1
US-09-620-405B-475
; Sequence 475, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620.405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-620-405B-475

Alignment Scores:
Pred. No.: 9,7e-94
Score: 917.00
Percent Similarity: 59.77%
Best Local Similarity: 48.94%
Query Match: 12.94%
DB: 4
Matches: 506
Conservative: 112
Mismatches: 285
Indels: 145
Gaps: 83

Length: 1002
Matches: 506
Conservative: 112
Mismatches: 285
Indels: 145
Gaps: 83


```
QY 2992 ATTGCATGTTGAAAAGGAATTCGCCA---TGCAAACTGGAATAGCCACACTA---AACA 3045
Db 709 snCyMetLeuLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHis- 728
QY 3046 CCAATCCAGAAAGAAATAAATATC-TTGTAGGACTTAAGATTTTAAAGA---AAGAT 3101
Db 729 GlnTyGlnGluLysGluAsnLysTyPheGluAspIleLysIleLeuLysGluLysAsn 748
QY 3102 GCTGACACTAG-ATGACCCCTAAACTGAAAGGAATCATTACTAAAG---GGATCTCAA 3157
Db 749 AlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGln 768
QY 3158 TATGTGGGAGCT---AAAGTTCTGATAGCTGAGAAACAATGCTCATCTAAATTCAGA 3214
Db 769 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 788
QY 3215 AAAACAGACAAGAAAA-----CTAGAGCGAGAATTCGAATA-CACCATCTGACTGCG 3267
Db 788 ysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisPro----- 805
QY 3268 TTCGCTGTACAAGC-----CATGATCAATTTGTGACATCAAGAAAAAGTAAGA- 3314
Db 806 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 825
QY 3315 --ACCTGCTCCACATTGAG---GAGATCCTGTTTGCAAGAAATGAATGTTGATGTG- 3368
Db 825 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 845
QY 3369 --AGACTGACGATTAACCAATCAGTGTCTCCATCACCCTTTC---TGAGCTCAAAAGAAA 3423
Db 845 erSerThrIleTy-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 864
QY 3424 TCCAAACCTCAAAATTAATCT---CAATTAGCAGGAGATGCTTAAGAGA---AATACATT 3477
Db 865 SerLysSerLeuLysIleAsnLeuAsnTyAlaGlyAsp-AlaLeuATGluAsnThrIle 884
QY 3478 GGTTCAGACATCAACAAGAGA---CAACGTGAACAACAGTGTCAAT---GAAGGAGCG 3531
Db 884 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 904
QY 3532 AACACATGTATCAAAAC-----GACAAGATAATGGAACAAACACCTGAAACAGAGATC 3585
Db 904 LuHisMetTyGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerL 924
QY 3586 TCTAGTCAGAAATATTTCACACTACAGCAAAAATTTGCGCTTCACAGCAATT---AGT 3641
Db 924 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 943
QY 3642 CATGCACATAGAAAGCTGACACAAAAGCAGATAACAAT-----GATATT 3686
Db 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
QY 3687 CATTTCTTG---AGAGAAATGCAACATCTCTTAAAGAGAAAAGAGGAGATATT--- 3741
Db 960 HisPheLeuGluArgLysMetGlnHisHisLeuLeuLysGluLysAsnGlu-GluIlePh 979
QY 3742 -AATTACAATACCATTTTAAACCGGTATTT---CAATATGAAAAAGAGAAA 3789
Db 979 eAsnTyAsnAsnHisLeuLysAsnArgIleTyGlnTyGlnLysGluLys 996
```

RESULT 2

```
US-09-604-287A-475
; Sequence 475, Application US/09604287A
; Patent No. 6585572
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuguo
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
```

```
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-604-287A-475
Alignment Scores:
Pred. No.: 9,7e-94 Length: 1002
Score: 917,00 Matches: 506
Percent Similarity: 59,77% Conservative: 112
Best Local Similarity: 48,94% Mismatches: 285
Query Match: 12,94% Indels: 145
DB: 4 Gaps: 83
US-09-602-362E-22 (1-4115) x US-09-604-287A-475 (1-1002)
```

```
QY 1031 ATGAGTCCCGAAAGAACAACACTGAGAA--ATTACGTGGCGAGA-AAAGGAAGACCTAGG 1086
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrIrpAlaAlaLysGlyArgProArg 20
QY 1087 AAGAC-GCATGGGAGAAAAAGAAA---CCCTGTAAGATGATGTCGT-GCAAGAGTAACA 1141
Db 21 LysIleAlaLrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY 1142 TCTAATAAATAAAG---TTTTGAAAAGAAATCTAAGATGTT-GCATCTCCACAAAG 1197
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
QY 1198 ---AATCATCTACAAACAGTGCATGATCAGAGG---TCCCATCAGATCCAAACAAG 1251
Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
QY 1252 GAAGATGAAGAATATCTTTT-GATCTCGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
Db 81 GluAspGluGluTySerCysAspSerArg---SerLeuPheGluSerSerAlaLysIle 99
QY 1305 CAAGTG---GTATACCTGAGTCTATATATAAAAGTAATAGATAAATAAGAG---TAG 1358
Db 100 GlnValCysIleProGluSerIleTyGlnLysValMetGluIleAsnArgGluValGlu 119
QY 1359 AAAGCTCTCTAAAAGCCATCTGCTTCAAGCCGCAATTGAAAGCAAAA---CTCTTTCCA 1415
Db 120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
QY 1416 AATAAGCCCTTTG-----AATGAAGAATGAACAACATTGAAGCAGATCCCTG-TTC 1465
Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
QY 1466 CCACCG---AATCCAAACAAAGAGACTATAAGAAATTTCTTGGGATTTCTAGAGTCTCTG-- 1520
Db 158 ProProGluSerLysGlnLysAspTyGluGluAsnSerTrpAspSer-GluSerLeuCy 177
QY 1521 -GAGACTGTTTACAGAAG---ATTGTGTTTACCAGGCTACACATCAAAAAGAAATAGAT 1576
Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
QY 1577 AAAAA---AATGGAAAA--TAGAAGAGTCCCTAATAAAGTGGTCTTCTGTAAGGCTACCT 1630
Db 197 sPLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
QY 1631 CGGAAT---GAATTTCTTATTCACATAAAGCCT---AGAATTGAAGACATGCCAACTTTC 1684
Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
```


Db 944 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 959
 QY 3687 CATTCTTGTG--AGAGGAAATGCAACACATCTCTTAAAGAGAAAGAGAGATATT-- 3741
 Db 960 HisPheLeuGluArgLysMetGlnHisLysLeuLeuLysGluLysAsnGlu-GluIlePh 979
 QY 3742 -AATTACATACCAATTAAACACCGTATATT--CAATATGAAAAAGAGAAA 3789
 Db 979 eAsnIyrAsnAsnHisLysLysAsnArgIleIleTyrGlnIyrGluLysGluLys 996

RESULT 3

US-09-834-759-475
 ; Sequence 475, Application US/09834759
 ; Patent No. 6680157
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Dillon, David C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepier, William T.
 ; APPLICANT: Henderson, Robert A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 ; FILE REFERENCE: 210121.470C9
 ; CURRENT APPLICATION NUMBER: US/09/834,759
 ; CURRENT FILING DATE: 2001-04-13
 ; NUMBER OF SEQ ID NOS: 547
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 475
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(1002)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-834-759-475

Alignment Scores:
 Pred. No.: 9,7e-94 Length: 1002
 Score: 917.00 Matches: 506
 Percent Similarity: 59.77% Conservative: 112
 Best Local Similarity: 48.94% Mismatches: 285
 Query Match: 12.94% Indels: 145
 DB: 83 Gaps: 83

US-09-602-362E-22 (1-4115) x US-09-834-759-475 (1-1002)

QY 1031 ATGAGTCCCGAAAGAAACACTGAGAA---ATTACGTGGGCGA-AAAGGAGACCTAGG 1086
 Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 20
 QY 1087 AAGAC-GCATGGGAGAAAAAGAA---CCCTGTAAAGATGGATCGCT-GCAAGAGTAACA 1141
 Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
 QY 1142 TCTATAAATAAG--TTTGTAAAGAAATCTAAGATGT-CCATGTCACAAAG 1197
 Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
 QY 1198 ---AATCATCTCAAGCAAGTCCCATGATCAGAG--TCCCATCAGATCCAAACAAG 1251
 Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
 QY 1252 GAAGATGAAGAATATCTTTT-GATTCTCGGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
 Db 81 GluAspGluLysThrSerCysAspSerArg---SerLeuPheGluSerAlaLysIle 99
 QY 1305 CAAGTG---GTATACCTGAGCTCTATATAAAGTAATGAGATAAAGAG---TAG 1358
 Db 100 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 119

QY 1359 AAAGCCCTCTAAAGCCATCTGCTTCAAGCCGCCATTGAAAGCAAAA---CTCTTTTCCA 1415
 Db 120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
 QY 1416 AATAAGCCTTTG-----AATGAAGAATGAACAAACATTAAGCAGATCCGTG-TTC 1465
 Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
 QY 1466 CCACCG---AATCCAAACAAAGGACTATAGAAAATTTCTTGGGATTTCTAGAGTCTCTG-- 1520
 Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
 QY 1521 -GAGACTGTTTACAGAAG---ATTGTGTTTACCAAGGCTTACACATCAAAAAGAAATAGAT 1576
 Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
 QY 1577 AAAAA---AATGGAAGA---TAGAAGAGTCCCTAATAAAGTGTCTTCTCAAGGCTACCT 1630
 Db 197 spLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
 QY 1631 CGGAAT---GAAATTTCTATTCCACTAAAGCCT---AGAATTGAAGACATGCAAACTTTC 1684
 Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
 QY 1685 AAAGC-GAGCCTCGGGAAG---CATCTGCTTCGAGCCTCCACTGAAATC---AAAAGT 1737
 Db 237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
 QY 1738 CTGTCCCAATTAAGCCTTGAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 1794
 Db 257 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 276
 QY 1795 CCATCAGAAATCCAAACAAAGGACTATAGAAA---TTCTGGGATATCTGAGTCTCTG-- 1849
 Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSer-GluSerLeuCy 296
 QY 1850 -TAGACTGTTTACAGAAGGATG-TGTTTACCAGG---CTGCGCTCAAAAAGAAATAGA 1904
 Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleAs 316
 QY 1905 TAAATTAATGAAATTAGAAGGCT---CCCTGTAAAGTGTCTTCTGAGGCTAACT-- 1959
 Db 316 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 336
 QY 1960 -GGGAATGAAGTCTTATCCAACTAAAGCTAGA---ATTGATGATGATGCAAACTTCAA 2015
 Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
 QY 2016 AG---CAGGCTCCCGAGAGCCATCTGCTTCGAGCCTCCATGAAA---TCAAAAGTCT 2069
 Db 356 ysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
 QY 2070 GTCCAAATAAAC---CTTGAATTAAGATGAACAAACATTAAGCAGATGAGAA---CT 2123
 Db 376 alProAsn-LysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 395
 QY 2124 CCCATCA---AATCCAAACAAAGGACTATGAGAAGTTCTTGGGATTTCTGAGTCTCTG- 2179
 Db 395 uproSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSer-GluSerLeuCy 415
 QY 2180 --TAGACTGTTTCCAGAGGATGTTG---TTTACCAGGCTTACACATCAAAAAGAAATAGA 2234
 Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleA 435
 QY 2235 TAAA---TAATGGAATAAGAGTCTCT-GATAATGAGGTTTCTGAAGGCTCCCT- 2289
 Db 435 spLysIleAsnGlyLysLeuGluGluSerProAsnAspGlyPheLeuLysAlaProC 455
 QY 2290 --GAGAATGAAGTCTTATCCACTAAAGCCT---TGAATTCATGGCATGCAAACTTTC 2344
 Db 455 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
 QY 2345 AAGAGGCTCCCG---AGAGCCATCTCTGAGCCTGCATTTGAAAT---GAAAAGTCT 2398


```
QY 3046 CCATCCAGGAAGGAATAATATC-TTTGAGGACTTAAGATTAAAGA---AAAGAT 3101
Db 822 GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysLeuLysGluLysAsn 841
QY 3102 GCTGAACCTTAG-ATGACCCCTAAACTGAAAAGGAATCATTAATAAAG---GGATCTCAA 3157
Db 842 AlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAlaSerGln 861
QY 3158 TATGTGGGAGCT---AAAGTTCTGATAGCTGAGAAACAATGCTCATTTAAATTGAGGA 3214
Db 862 TyrSerGly-GlnLeuLysValLeuIleAlaGlu-AsnThrMetLeuThrSerLysLeuL 891
QY 3215 AAAACACAGCAAGAAAA---CTAGAGGACAGAAATTGAATA-CACCATCCTGACTGGC 3267
Db 881 ySGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisPro----- 898
QY 3268 TTGCTGTACAAGC-----CATGATCAATTTGTGACATCAAGAAAAAGTAAGA- 3314
Db 899 --ArgLeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnG 918
QY 3315 --ACCTGCTTCCCATTTAG---GAGATGCTGTTTGCAGAGAAAATGAATGTTGATGTG- 3368
Db 918 luProAlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValS 938
QY 3369 --AGAGTACGATATTAAACAATGAGTGCTCCATCACCACCTTTC---TGAGCTCAAAGGAAA 3423
Db 938 erSerThrIleTyr-AsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLys 957
QY 3424 TCCAAAACCTTAATAATTTATCT---CAATTAGCAGGAGATGCTTAAGAGA---AATACATT 3477
Db 958 SerLysSerLeuLysIleAsnLeuAsnTyr-AlaGlyAsp-AlaLeuArgGluAsnThrIle 977
QY 3478 GGTTTTCAAGACATCACAAAGAGA---CAAGCTGAAAACAGTGTCAAT---GAAGGAAGCG 3531
Db 977 uValSerGluHis-AlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaG 997
QY 3532 AACCATGTATCAAAAC-----GACAAGATATGMAACAAACACCTGAACACAGAGTC 3585
Db 997 luHisMetTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerL 1017
QY 3586 TCTAGTCAGAAATATTTCACATACAGCAAAATTTGTGCTTCCACAGCAATT---AGT 3641
Db 1017 eu-AspGlnLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuVal 1036
QY 3642 CATGCAATAGAAAGCTGCAACAAAGACAGATAACAAT-----GATATT 3686
Db 1037 HisAlaHisLys-----LysAlaAspAsnLysSerLysIleThrIleAspIle 1052
QY 3687 CATTTCTTG---AGAGAAATGCAACACATCTCTTAAGAGAAAAGAGGAGATATT-- 3741
Db 1053 HisPheLeuGluArgLysMetGlnHisHisLeuLeuLysGluLysAsnGlu-GluIlePh 1072
QY 3742 -AATTACAATACCATTTTAAACCGTATATT---CAATATGAAAAAGAGAAA 3789
Db 1072 eAsnTyrAsnAsnHisLeuLysAsnArgIleTyrGlnTyrGlnLysGluLys 1089
```

RESULT 5

```
US-09-834-759-493
; Sequence 493, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugui
; APPLICANT: Dillion, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
```

```
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(1095)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-834-759-493
```

Alignment Scores:

Pred. No.:	1,94e-93	Length:	1095
Score:	914.50	Matches:	506
Percent Similarity:	59.77%	Conservative:	112
Best Local Similarity:	48.94%	Mismatches:	285
Query Match:	12.91%	Indels:	145
DB:	4	Gaps:	83

US-09-602-362E-22 (1-4115) x US-09-834-759-493 (1-1095)

```
QY 1031 ATGAGTCCCGAAAGAAACACTGAGAA---ATTACGTGGGCAGAGA-AAAGGAAGACCTAGG 1086
Db 94 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 113
QY 1087 AAGAC-GCATGGGAGAAAAGAAA---CCCTGTAAAGATGATCGCT-GCAAGAGTAACA 1141
Db 114 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 133
QY 1142 TCTATAAACTAAAG---TTTGTAAAGGAATCTAAGATGTT-GCATGTCACAAAAG 1197
Db 134 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 153
QY 1198 --AATCATCTACAAGCAAGTGCATGATCAGAGG--TCCCATCAGATCCAAACAAGG 1251
Db 154 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 173
QY 1252 GAAGATGAAGAATATCTTTT-GATTCGGGTCTCTTTG-----AAGTCTCGCAAGATT 1304
Db 174 GluAspGluGluTyrSerCysAspSerArg---SerLeuPheGluSerSerAlaLysIle 192
QY 1305 CAAGTG---GTATACCTCAGTCTATATAAAGAGTAATGAGATAATAAGAG--TAG 1358
Db 193 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 212
QY 1359 AAAGCCTCTAAAAGCCATCTGCCCTTCAAGCGCCCATTTGAAGCAAAA---CTCTTTCCA 1415
Db 213 GluProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 232
QY 1416 AATAAGCCTTTG-----AATGAAGATGAACAACATTTGAAGCAGATCCGTG-TTC 1465
Db 233 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 250
QY 1466 CCACCG---AATCCAAAACAAGGACTATAGAAAATCTTTGGGATTTCTAGAGTCTCTG-- 1520
Db 251 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 270
QY 1521 -GAGAGCTTTTACAGAAG---ATTGTGTTTACCAAGCTACACATCAAAAAGAAATAGAT 1576
Db 270 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 290
QY 1577 AAAAA---AATGGAAAA---TAGAAGAGTCCCTAATAAAGTGGTCTTCTCAAGGCTACCT 1630
Db 290 spLysIleAsnGlnLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 310
QY 1631 CGGAAT---GAAATTTCTATTCCACTAAAGCCT---AGAATTGAAGACATGCAAACTTTC 1684
Db 310 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 329
QY 1685 AAAGC-GAGCCTCCGGAAGC---CATCTGCTTCGAGCCCTCCACTGAAATC---AAAAGT 1737
```



```
QY 2399 GTTCCAAATAAAC---CTTGGAAATTAAGATGAACAAACATTGTGAGACAGATCAGAT----- 2450
Db 495 ValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArgAlaAspGlnVal 513
QY 2451 -TTCCCTTCAGATCAAAACAAAGAGGTTTGACAAA---TTCTGGGATTCTGAAGTC- 2505
Db 513 tPheProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGlu-SerL 533
QY 2506 --CCGTGGACTGTTTCACAGAAGAGTGTGTGTACC---CAGCTACACACAAAAGAAA 2560
Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
QY 2561 GGATAAAAT--AGTGAAATATAGAGATTCTACTAGCCTATCAAAAT--CTTGGTACA 2614
Db 553 MetAspLysLysSerGlnLysLeuGluAspSerThrSerLeuSerLysLysLeuAspThr 572
QY 2615 GTTCAT---CTTGTGAAGGCAAGGGAACCTTCAAAAAGACACTGTGAACA-CGTACAGCA 2670
Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
QY 2671 AAATGGACAAAT--GAAAAGAGTTTGTGTACTGAAAAAGAAC---TGTGAGAACAAA 2724
Db 593 LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 612
QY 2725 AGAAATAAATCACAGTAGAGAA---CCAAAAGTTAAATGGCAACAGAG 2770
Db 613 GluLeuLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGlu 628

RESULT 7
US-09-433-826B-469
; Sequence 469, Application US/09433826B
; Patent No. 6579973
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C4
; CURRENT APPLICATION NUMBER: US/09/433,826B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 474
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 469
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (310)
; NAME/KEY: Xaa = Any Amino Acid<221> unsure
; LOCATION: (429)
; NAME/KEY: Xaa = Any Amino Acid<221> unsure
; LOCATION: (522)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-433-826B-469

Alignment Scores:
Pred. No.: 2,65e-56 Length: 650
Score: 583.50 Matches: 317
Percent Similarity: 60.47% Conservative: 73
Best Local Similarity: 49.15% Mismatches: 185
Query Match: 8.23% Indels: 80
DB: 4 Gaps: 52

US-09-602-362B-22 (1-4115) x US-09-433-826B-469 (1-650)
QY 1031 ATGAGTCCCGAAAGAAACACTGAGAA---ATTACGTGGCAGCA-AAAGAGACCTAGG 1086
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 20
```

```
QY 1087 AAGAC-GCATGGGAGAAAAAGAAA---CCCTGTAAAGATGGATCGGT-GCAAGAGTAACA 1141
Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY 1142 TCTAATAAATAAAG---TTTGTGAAAAAGAGAAATCTAAGATGTT-GCATGTCCCAAAAG 1197
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
QY 1198 ---AATCATCTACAAAGCAAGTCATGATCAGAGG---TCCCATCAGATCCAAACAAGG 1251
Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
QY 1252 GAAGATGAAGAAATATCTTT-GATTCTCGGCTCTCTTG-----AAGTCTCGAAGATT 1304
Db 81 GluAspGluGluTyrSerCysAspSerArg---SerLeuPheGluSerSerAlaLysIle 99
QY 1305 CAAGTG---GTATCATCTGAGTCTATATATAAAAGTAAATAGATATAATAAGAAAG---TAG 1358
Db 100 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 119
QY 1359 AAAGCCTCTAAAAGCCATCTCGCTTCAAGCGCCATTTGAAAGCAAAA---CTCTTCCCA 1415
Db 120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
QY 1416 AATAAGCCTTTTG-----AATCAAGAATGAACAAACATTGAAGCAGATCCGTG-TTC 1465
Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
QY 1466 CCACCG---AATCCAAACAAAGGACTATATAAGAAATTTCTGGGATTTAGAGTCTCTG-- 1520
Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
QY 1521 -GAGACTCTTTACAGAAGG---ATTGTGTTTACCAAGGCTACACATCAAAAGAAATAGAT 1576
Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
QY 1577 AAAAA---AATGGAATA---TAGAAGACTCCTTAATAAGTGTCTTCTGAAGCGCTAGCT 1630
Db 197 sPylleAsnGlyLysLeuGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
QY 1631 CGGAAT---GAAATTTCTATTCCTAAAGCCT---AGAATTGAAGACATCAAACTTTC 1684
Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
QY 1685 AAAGC-GAGCCTCCGGGAAGC---CATCTGTTTCGAGCCTCCACTGAAATC---AAAAGT 1737
Db 237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
QY 1738 CTGTCCCAATAAGCCTTGGAAATGAAATGAAACATGAGGCGCAGATGAGAA---CTC 1794
Db 257 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluLeu 276
QY 1795 CCATCAGAAATCCAAACAAAGGACTATAGAATA---TTCTGGGATCTGGAGTCTCTG-- 1849
Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 296
QY 1850 -TAGACTGTTTCACAGAAGAGTTG-TGTTTACCAAGG---CTGGCTCAAAAAGAAATAGA 1904
Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleAs 316
QY 1905 TAAATAATGGAATAATTAGAAGGTT---CCCTGTTTAAAGTGGTCTCTTCGAGGCTAACT-- 1959
Db 316 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 336
QY 1960 -GGGAATGAAAGTTCTATTCCAACCTAAAGCCTAGA---ATTGATGACATGCAAACTCAA 2015
Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
QY 2016 AG---CAGGCTCCCGAGAGCCATCTGCTTCGAGCCTCCATTGAAA---TCAAAAGTCT 2069
Db 356 ysAlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
QY 2070 GTCCAAATAAAC---CTTGGAAATTAAGAAATGAACAAACATTTGAGGCGAGATGAGAA---CT 2123
```

```

Db      376  alProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 395
QY      2124  CCCATCA---AATCCAACAAGCACTATGAGAAAGTCTTGGGATTCCTGAGTCTCTG-- 2179
Db      395  uProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuC 415
QY      2180  --TAGACTGTTTCCAGAGGATGTG---TTTACCAGGCTACACATCAAAAGAAATAGA 2234
Db      415  ysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleA 435
QY      2235  TAAAA---TAATGAAAATAGAGAGTCTCT-GATAATGAGTGTTCGAGGCTCCCT- 2289
Db      435  spLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProC 455
QY      2290  --GAGAAATGAAGTCTTATCCACTAAAGCCT--TGAATGTGCGCATCAAACTTTCA 2344
Db      455  ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
QY      2345  AAGCAGGCTCCCG---AGAGCCATCTGCTTCGAGCCTGCATTGAAAT---GAAAAGTCT 2398
Db      475  ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
QY      2399  GTTCCAAATAAAC---CTTGGAAATTAAGATGAACACATTGAGACAGATCAGAT---- 2450
Db      495  ValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
QY      2451  -TTCCTTCAAGATCAAAACAAGAGTTTCAGAAAA---TTCTGGGATTCGAAGTCT- 2505
Db      513  tPheProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGlu-SerL 533
QY      2506  --CCGTGAGCTGTTTCACAGAGGATGCTGTGTACC---CAGGCTACACACAAAAAGAA 2560
Db      533  euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
QY      2561  GGATAAAAT---AGTGAAATTAAGAGATTCAGCTTATCAAAAT---CTTGTGTACA 2614
Db      553  MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
QY      2615  GTTCAT---CTTGTGAAGCAGGAGGAAGTTCAAAAGACATCTGTGACAC-CGTACAGGA 2670
Db      573  ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
QY      2671  AATGGAAACAAT---GAAAAGAGTTTGTGTACTGAAAAGAAC---TGTCAAGACAAA 2724
Db      593  LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysLysLeuSerGluAlaLys 612
QY      2725  AGAAATAATACAGTAGAGAA---CCAAAAGTTAAATGGGAACAGAG 2770
Db      613  GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGlu 628

```

RESULT 8

```

US-09-604-287A-469
; Sequence 469, Application US/09604287A
; Patent No. 6586572
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604.287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 469
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (310)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: unsure
; LOCATION: (429)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: unsure
; LOCATION: (522)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-604-287A-469

```

Alignment Scores:

```

Pred. No.: 2,65e-56 Length: 650
Score: 583.50 Matches: 317
Percent Similarity: 60.47% Conservative: 73
Best Local Similarity: 49.15% Mismatches: 185
Query Match: 8.23% Indels: 80
DB: 4 Gaps: 52

```

US-09-602-362E-22 (1-4115) x US-09-604-287A-469 (1-650)

```

QY      1031  ATGAGTCCCGAAAGAAAGAACTAGAGAA---ATTACGTGGGCAGAG-AAAGGAAGACCTAGG 1086
Db      1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 20
QY      1087  AAGAC-GCATGGGAGAAAAGAAA---CCCTGTAAGATGGATCGGT-GCAGAGATGAACA 1141
Db      21 LysIleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY      1142  TCTATAAACTAAAG---TTTTGAAAAGAGAACTAAGATGTT-GCATGTCCACAAAAG 1197
Db      41 SerAsnLysThrLysValLeuLysGlyArgSerLysMetIleAlaCysProThrLys 60
QY      1198  ---AATCATCTACAAGCAAGTCCATGATCAGAGG---TCCCATCAGATCCCAACAAAG 1251
Db      61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 80
QY      1252  GAAGATGAAGATATCTTTT-GATTCTCGGTCTCTTTG-----AAGTTCTGCAAGATT 1304
Db      81 GluAspGluGluTyrSerCysAspSerArg---SerLeupheGluSerSerAlaLysIle 99
QY      1305  CAAGTG---GTATACCTGAGTCTATATAAAAAAGTAATAGATAATAATAAGAG---TAG 1358
Db      100 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 119
QY      1359  AAGCCTCTAAAGCCCATCTGCTTCAAGCCGCCATTGAAAGCAAAA---CTCTTTTCCA 1415
Db      120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
QY      1416  AATAGCCCTTTTG-----AATCAAGAATGAACAAACATTTGAAGCAGATCCCTG-TTC 1465
Db      140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
QY      1466  CCACCG---AATCCAACAAGGACTATAGAAAATTTCTGGGATTTAGAGTCTCTG-- 1520
Db      158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 177
QY      1521  -GAGACTGTTTACAGAGG---ATTGTGTTTACCAGGCTACACATCAAAAAGAAATAGAT 1576
Db      177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
QY      1577  AAAAA---AATGGAAGAA---TAGAAGAGTCCCTAATAAAGTGTCTTCTGAAGCTACCT 1630
Db      197 spLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLysAlaThrC 217
QY      1631  CGGAAT---GAAATTTCTATTCCACTAAAGCCT---AGAAATTGAAGACATGCAAACTTTC 1684
Db      217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
QY      1685  AAAGC-GAGCCTCCGGGAAGC---CATCTGTTTCGAGCCTCCACTGAATC---AAAAGT 1737
Db      237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256

```

QY	1738	CTGTCCAAATAGCCTTGAATGAAATAAATGAAACATGAGCGAGATGAGAA---CTC	1794
Db	257	VaiProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu	276
QY	1795	CCATCAGAATCCAAACAAAGGCATATAAGAAA--TTCTGGGATACTGAGTCTCTG--	1849
Db	277	ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerIeuCy	296
QY	1850	-TAGACTGTTTCACAGAAGGATTG-TGTTTACCAGG---CTCGGCTCAAAAAGAATAGA	1904
Db	296	sGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleAs	316
QY	1905	TAAATAATCGAAAATTAGAGGGT---CCCTGTTAAAGTGCTTCTGAGGCTAACT--	1959
Db	316	pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy	336
QY	1960	-GGGAATGAAAGTTCTATTCCAATAAGGCTAGA--ATTGATGACATCCAAACTCAA	2015
Db	336	sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL	356
QY	2016	AG---CAGGCGCTCCGAGAGCCATCTCCCTTCGAGCGCTCCATTGAA--TCAAAAGTCT	2069
Db	356	ysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV	376
QY	2070	GTCCAAATAAAC---CTTGGAAATTAGAATGAACAAACATTTAGGCGAGATGAGAA---CT	2123
Db	376	alProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe	395
QY	2124	CCCATCA--AATCCAAACAAGGACTATGAGAAAGTCTTGGGATCTCTGGAGTCTCG	2179
Db	395	uProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuC	415
QY	2180	--TAGACTGTTTCAGAAAGGATGGTG---TTTACCAGGCTACACATCAAAAAGAATAGA	2234
Db	415	ysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleA	435
QY	2235	TAAAA---TAATGGAAATAGAAGTCTCT-GNTATAGAGTTTCTGAGGCTCCCT-2289	
Db	435	spLysIleAsnGlyLysLeuGluGluSerProAsnAspGlyPheLeuLysAlaProC	455
QY	2290	--GAGAAATGAAAGTTCTATTCCACTAAAGGCT--TCGAATTGATGGCATGCAAACTTCA	2344
Db	455	ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL	475
QY	2345	AGCAGCGCTCCCG---AGAGCCATCTGCTTCGAGCTCGATTGAAAT---GAAAAGTCT	2398
Db	475	ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer	494
QY	2399	GTTCCAATAAAC---CTTGGAAATTAGAATGAACAAACATTTGACAGATCAGAT---	2450
Db	495	VaiProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr--LeuArgAlaAspGlnMe	513
QY	2451	-TTCCCTTCAGAAATCAAAACAAGAAGTTGAGAAA--TTCTGGGATCTCGAAGTCT-	2505
Db	513	tPheProSerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGlu-SerL	533
QY	2506	--CCGTGGACTGTTTCAGAAGGATGGTGTGTACC--CAGGCTACACACAAAAAGAAA	2560
Db	533	euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu	552
QY	2561	GGATAAAT---AGTGGAAATTAGAGATTCTACTAGCCTATCAAAAT---CTTGGTACA	2614
Db	553	MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr	572
QY	2615	GTTCAT--CTTGTGAAAGGCAAGGGAACCTTCAAAAAGACACTGTGAACA-CGTACAGGA	2670
Db	573	ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly	592
QY	2671	AAATGGAACAAT--GAAAAAGAGTTTGTGTACTGAAAAAGAAC---TGTCAACAACAA	2724
Db	593	LysMetGluGlnMetLysLysLysPheCysValLeuLysLysLysLysLeuSerGluAlaLys	612

```

QY 2725 AGAATAAATACAGTAGAGAA---CCAAAGTAAATGGGAACAAG 2777
Db 613 GluLeuYsserGln-LeuGluAsnGlnYsValYsTrpGluGlnGlu 628

RESULT 9
US-09-834-759-469
; Sequence 469, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 469
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (310)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: unsure
; LOCATION: (429)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: unsure
; LOCATION: (522)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-834-759-469

Alignment Scores:
Pred. No.: 2,65e-56 Length: 650
Score: 583.50 Matches: 317
Percent Similarity: 60.47% Conservative: 73
Best Local Similarity: 49.15% Mismatches: 185
Query Match: 8.23% Indels: 80
DB: 4 Gaps: 52

US-09-602-362B-22 (1-4115) x US-09-834-759-469 (1-650)

QY 1031 ATGAGTCCCCAAAGAAACACTCAGAA---ATTACGTGGGCAGAAAGGAAGA 650
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTIPAlaAlaLysGlyArg 650

QY 1087 AAGAC-CCATGGGAGAAAAGAAA---CCCTGTAAAGATGGATCGCT-GCAAGA 650
Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArg 650

QY 1142 TCTAATAAACTAAG---TTTCAAAAAGAAATCTAAGATGTT-CCATGTCCA 650
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysPro 650

QY 1198 ---AATCATCTACAAAGCAGTGCATGATCAGAG---TCCCATCAGATCCA 650
Db 61 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLys 650

QY 1252 GAAGATCAAGAAATATCTTT-GATTCTCGGGTCTCTTIG-----AAGTTCTG 650
Db 81 GluAspGluGluTrpSerCysAspSerArg---SerLeuPheGluSerSerAla 650

QY 1305 CAAGTG---GTATACCTGAGTCTATATATATAAAGTAATGAGTAATAAGAG 650
Db 100 GlnValCysIleProGluSerIleTrpGlnLysValMetGluIleAsnArgGlu 650

```

QY 1359 AAAGCTCTCTAAAGCCATCTGCTTCAAGCGCCATTGAAAGCAAAA---CTCTTTCCA 1415
Db 120 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 139
QY 1416 AATAAGCCTTTG-----AATCAAGAATGAACAACATTGAAGCAGATCCGTG-TTC 1465
Db 140 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 157
QY 1466 CCACGG---AATCAACAAGAGCTATAAGAAATCTTGGGATCTTAGAGTCTCTG-- 1520
Db 158 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSerLeuCy 177
QY 1521 -GAGACTGTTTACAGAAGG---ATTGTGTTTACCAAGGCTACACATCAAAAAGAATAGAT 1576
Db 177 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 197
QY 1577 AAAA---AATGGA---TAGAAGAGTCCCTAATAAAGTGGTCTTCTGAAGCTACCT 1630
Db 197 sPlysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 217
QY 1631 CGGAAT---GAAATTTCTATCCACTAAAGCCT---AGAATTGAAGCATGCAAACTTTC 1684
Db 217 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 236
QY 1695 AAAGC-GAGCCTCCGGGAGC---CATCTGCTTCGAGCCTCCACTGAATC---AAAGT 1737
Db 237 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 256
QY 1738 CTGTCCCAATAAGCCTTGAATCAAAATCAAAACATGAGGCAGATGAGAA---CTC 1794
Db 257 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 276
QY 1795 CCATCAGAAATCCAAACAAAGGACTATAAGAAA---TTCTGGGATCTGAGTCTCTG-- 1849
Db 277 ProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSerGluSerLeuCy 296
QY 1850 -TAGACTGTTTACAGAAGGATTG-TGTTTACCAAGG---CTGCGCTCAAAAGATAGA 1904
Db 296 sGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleAs 316
QY 1905 TAAATAATGAAAATAGAGGCT---CCCTGTTAAAGTGGTCTTCTGAGGCTAACT-- 1959
Db 316 PlysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCy 336
QY 1960 -GGGAATGAAGTCTTATTCACATCAAGCCTAGA---ATTGATGACATCAAACTCAA 2015
Db 336 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 356
QY 2016 AG---CAGGCTCCCGAGAGCATCTGCTTCGAGCCTCCATTGAAA---TCAAAAGTCT 2069
Db 356 ysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 376
QY 2070 GTCCAAATAAAC---CTTGGAAATTAAGAATGAACAAACATTGAGCGCAGATGAGAA---CT 2123
Db 376 alProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 395
QY 2124 CCCATCA---AATCCAAACRAAGGACTATGAGAAAGTCTTGGGATCTCGAGTCTCTG- 2179
Db 395 uProSerGluSerLysGlnLysAspTyrGluGluSerTrpAspSerGluSerLeuC 415
QY 2180 --TAGACTGTTTCCAGAGGATGGTG---TTTACCCAGGCTTACACATCAAAAAGATAGA 2234
Db 415 ysGluThrValSerGlnLysAspValCysLeuProLysAla---HisGlnLysGluIleA 435
QY 2235 TAAA---TAATGGAATAAGAGTCTCT-GATAATGAGTGTCTTCTGAAGGCTCCCT- 2289
Db 435 sPlysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProC 455
QY 2290 --GAGAATGAAGTCTTATTCACATAAGCCT---TGAATGTGATGGCATCAAACTTCA 2344
Db 455 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 475
QY 2345 AAGCAGGCTCCCG---AGAGCCTCTGCTTCGAGCCTGCATTGAAT---GAAAGTCT 2398

Db 475 ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 494
QY 2399 GTTCCAAATAAAC---CTTGGAAATTAAGAATGAACAAACATTGACACAGATCAGAT----- 2450
Db 495 ValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArgAlaAspGlnMe 513
QY 2451 -TTCCCTTCAGAAATCAAAACAAAGAGGTTGAGAAA---TTCTGGGATTTCTGAAGTCT- 2505
Db 513 tPheProSerGluSerLysGlnLys---ValGluGluAsnSerTrpAspSerGlu-SerL 533
QY 2506 --CGTGGAGCTTTTCACAGAAGGATGTGTGTACC---CAGGCTACACACAAAAGAAA 2560
Db 533 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 552
QY 2561 GGATAAAAT---AGTGGAAAATTAAGAACTTACAGATTCACCTAGCTATCAAAAT---CTTGGTACA 2614
Db 553 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 572
QY 2615 GTTCAT---CTTGTGAAGGCAAGGAACTTCAAAAGACACTGTGAACA-CGTACAGGA 2670
Db 573 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 592
QY 2671 AAATCGAACAAAT---GAAAAGAGTTTGTGTACTGTAAGAAAAGAAC---TGTCAAGACAAA 2724
Db 593 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 612
QY 2725 AGAAATAAATCACAGTAGAGAA---CCAAAAGTTAAATGGGAAACAAAGAG 2770
Db 613 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGlu 628

RESULT 10
US-09-620-405B-494
; Sequence 494, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 494
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(743)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-620-405B-494

Alignment Scores:
Pred. No.: 4, 2e-56 Length: 743
Score: 582.00 Matches: 317
Percent Similarity: 60.47% Conservative: 73
Best Local Similarity: 49.15% Mismatches: 185
Query Match: 8.21% Indels: 80
Gaps: 52
DB:

US-09-602-362E-22 (1-4115) x US-09-620-405B-494 (1-743)

QY 1031 ATGAGTCCCGAAAGAAACACTAGAA---ATTAGTGGGAGGA-AAAGGAGAGCTAGG 1086
Db 94 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 113

1087 AAGAC-GCATGGGAGAAAAAGAA---CCCTGTAAAGATGGATCGCT-GCAAGAGTAACA 1141
114 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaIleValThr 133
1142 TCATAATAAATAAG---TTTGAAGAAAGAAATTAAGATGTT-GCATGTCCACAAAG 1197
134 SerAsnLysThrLysValLysGluLysGlyArgSerLysMetIleAlaCysProThrLys 153
1198 ---AATCATCTCAAGACAGTGCATGATCAGAGG---TCCCATCAGATCCAAACAGG 1251
154 GluSerSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 173
1252 GAAGATGAAGATATTCTTT-GATTCTCGGGTCTCTTTG-----AAGTCTCGAAGATT 1304
174 GluAspGluGlnLysSerCysAspSerArg---SerLeuPheGluSerSerAlaLysIle 192
1305 CAAGTG---GTATACCTGAGTCTATATATAAAGTAATAGATAAATAAGAG---TAG 1358
193 GlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArgGluValGlu 212
1359 AAGAGCTCTCAAGAGCCATCGCTTCAAGCCGCCATCGAAGCAAAA---CTCTTTTCCA 1415
213 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 232
1416 AATAAGCCTTTG-----AATGAAGATGAACAAACATCGAACGATCGGTG-TTC 1465
233 AsnLysAlaPheGluLysLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 250
1466 CCACCG---AATCCAAACAAAGGACTATAGAATAATCTTGGGATCTAGAGTCTCTG-- 1520
251 ProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSer-GluSerLeuCy 270
1521 -GAGACTGTTTACAGAGG---ATTGTGTTTCAAGGCTPACACATCAAAAGAAATAGAT 1576
270 sGluThrValSerGlnLysAspValCysLeuProlLysAlaThrHisGlnLysGlu-IleA 290
1577 AAAAA---ANTGAAAA---TAGAGAGTCCCTTAATAAGTGTCTCTGAGGCTACCT 1630
290 sPlysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLysAlaThrC 310
1631 CGGAAT---GAAATTTCTATTCACCTAAAGCCT---AGAATTTGAAGACATGCAAACTTTC 1684
310 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThrPhe 329
1685 AAGC-GAGCTCGGGAAGC---CATCTGCTCGAGCTCCACTGAATC---AAAAGT 1737
330 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
1738 CTGTCCCAATAAGCCTTGAATGAAATGAAACATCGAGGCAGATCAGAA---CTC 1794
350 ValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 369
1795 CCATCAGAAATCCAAACAAAGGACTATAGAATAA---TTCTGGGATCTGAGTCTCTG-- 1849
370 ProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuCy 389
1850 -TAGACTGTTTACAGAGGATTC-TGTTTACCAAGG---CTGGCTCAAAAAGATAGA 1904
389 sGluThrValSerGlnLysAspValCysLeuProlLysAla---HisGlnLysGluIleAs 409
1905 TAAATTAATGGAATAATTAGAGGT---CCCTGTTTAAAGTGTCTCTGAGGCTAACT-- 1959
409 pLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLysAlaAsnCy 429
1960 -GGAAATGAAGTTCATTCCAACTAAAGCTAGA---ATTGATGATCATGCAAACTTCAA 2015
429 sGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMetGlnThrPheL 449
2016 AG---CAGGCTCCGAGGACCATCTCCCTTCAGAGCTCCATTCAGAA---TCAAAAAGTCT 2069
449 ysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerV 469
2070 GTCCAAATAAAC---CTTGGAAATTAAGATGAACAAACATGTAGGCGCAGATGAGAA---CT 2123

469 alProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp-GluIleLe 488
2124 CCCATCA---AATCCAAACAAAGGACTATGAGAAGTTCTTGGGATCTCTGAGTCTCTG- 2179
488 uProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer-GluSerLeuC 508
2180 --TAGACTGTTTCCAGAGGATCGGTG---TTTACCAGGCTACACATCAAAAAGAAATAGA 2234
508 ysGluThrValSerGlnLysAspValCysLeuProlLysAla---HisGlnLysGluIleA 528
2235 TAAAA---TAATGGAATAATAGAGTCTCT-GATAATGAGGTTTCTGAGGCTCCCT- 2289
528 sPlysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProC 548
2290 --GAGAATGAAGTCTTATTCACCTAAAGCCT---TGAATTCATGGCATGCAAACTTTCA 2344
548 ysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheL 568
2345 AAGCAGGCTCCCG---AGAGCCATCTCTGAGGCTGCTGAAAT---GAAAAGTCT 2398
568 ysAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluMetGlnLysSer 587
2399 GTTCCAAATAAAC---CTTGAATTAAGATGAACACATTCAGACAGATCAGAT----- 2450
588 ValProAsn-LysAlaLeuGluLysAsnGluGlnThr---LeuArgAlaAspGlnMe 606
2451 -TTCCCTTCAGAAATCAAAACAAAGAGGTTGAGAAA---TTCTGGGATTCGAGTCT- 2505
606 tPheProSerGluSerLysGlnLys---ValGluGluAsnSerTrpAspSerGlu-SerL 626
2506 --CGGTGACTGTTTACAGAGGATGCTGTGTACC---CAGGCTACACAAAGAAAGAA 2560
626 euArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHisGlnLysGlu 645
2561 GGATAAAT---AGTGGAAAAATTAGAAATTCACCTAGCCTATCAAAAT---CTTGTGTACA 2614
646 MetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThr 665
2615 GTTCAT---CTTGTGAAGGCAAGGAACTTCAAAAGACACTGTGAACA-CGTACAGGA 2670
666 ValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGly 685
2671 AAATCGAACAAAT---GAAAAAGAGTTTGTCTACTGAAAAAGAAC---TGTACAGAACAA 2724
686 LysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuSerGluAlaLys 705
2725 AGAATAAATCACACTAGAGAA---CCAAAAGTTAAATGGAAACAAGAG 2770
706 GluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGlu 721

RESULT 11
US-09-834-759-494
; Sequence 494, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 494
; LENGTH: 743
; TYPE: PR1

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(743)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-834-759-494

Alignment Scores:
Pred. No.: 4,28-56 Length: 743
Score: 582.00 Matches: 317
Percent Similarity: 60.47% Conservative: 73
Best Local Similarity: 49.15% Mismatches: 185
Query Match: 8.21% Indels: 80
DB: 4 Gaps: 52

US-09-602-362E-22 (1-4115) x US-09-834-759-494 (1-743)
QY 1031 ATGAGTCCGAAAGAAACACTGAGAA---ATTACGTGGGCGAGA-AAAGGAAGACCTAGG 1086
DB 94 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 113
QY 1087 AAGAC-GCATCGGAGAAAAAGAAA---CCCTGTAAGATGATGCGT-GCAAGAGTAACA 1141
DB 114 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaLaa-gValThr 133
QY 1142 TCTAATAACTAAAG---TTTTGAAAGAGAAATCTAAGATGT-GCATGTCCACAAAAG 1197
DB 134 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 153
QY 1198 ---AATCATCTACAAAGCAAGTCCATGATCAGAGG---TCCCATCAGATCCAAACAGG 1251
DB 154 GluSerThrLysAlaSerAlaAsnAspGlnArgPheProSerGluSerLysGlnGlu 173
QY 1252 GAAGATGAAGATATCTTT-GATTCTCGGTCTCTTTG-----AAGTTCTCAGAGATT 1304
DB 174 GluAspGluLysSerCysAspSerArg---SerLeuPheGluSerAlaLysIle 192
QY 1305 CAAGTG---GPATACCTGAGTCTATATATAAAGTAAATGAGATAATAAGAAAG---TAG 1358
DB 193 GlnValCysIleProGluSerIleTyrglnLysValMetGluIleAsnArgGluValGlu 212
QY 1359 AAAGCCTCCTAAAGCCTCTGCTTCAAGCCGCTTGAAGCAAAA---CTCTTTCCA 1415
DB 213 GluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSerValPro 232
QY 1416 AATAAGCCTTTG-----AATGAAGATGAACAAACATTTGAAGCAGATCCGCTG-TTC 1465
DB 233 AsnLysAlaPheGluLeuLysAsnGluGln-----ThrLeuArgAlaAspProMetPhe 250
QY 1466 CCACCG---AATCAAACAAGAGCTATAGAAAATTTCTGGGATTTCTAGAGTCTCTG-- 1520
DB 251 ProProGluSerLysGlnLysAspTyrgluGluAsnSerTrpAspSer-GluSerLeuCy 270
QY 1521 -GAGACTGTTTACAAAGG---ATTGTGTTTACCAAGGCTACACATCAAAAAGAATAGAT 1576
DB 270 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu-IleA 290
QY 1577 AAAAA---ATGGAAGA---TAGAAGAGTCCCTAATAAAGTGTCTTCTGAAGGCTACCT 1630
DB 290 spLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrC 310
QY 1631 CGGAAT---GAAATTTCTATTCCACTAAAGCCT---AGAATTGAAGACATGCAAACTTTC 1684
DB 310 ys-GlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 329
QY 1685 AAAGC-GAGCCTCGGGAAGC---CAPTCGCTTCGAGCCTCCACTGAAATC---AAAAAT 1737
DB 330 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
QY 1738 CTGTCCCAATAAGCCTTGGATGAAATGAAACATCGAGGCAGATGAGAA---CTC 1794
DB 350 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeu 369

```



```

US-09-620-405B-473
; Sequence 473, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 473
; LENGTH: 445
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-620-405B-473

Alignment Scores:
Pred. No.: 3-72e-38 Length: 445
Score: 421.50 Matches: 226
Percent Similarity: 61.49% Conservative: 55
Best Local Similarity: 49.45% Mismatches: 119
Query Match: 5.95% Indels: 60
DB: 4 Gaps: 38

US-09-602-362e-22 (1-4115) X US-09-620-405B-473 (1-445)
QY 1894 AAAAGTAATGATGATAAATGGAATGAGAGGT---CCCTGTTAAAGTGTCTCTG 1950
Db 1 LysGluIleAspLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeu 20
QY 1951 AGCTAACT---GGGAATGAAGTCTATTCCAACTAAAGCCTPAGA---ATTGATGACAT 2004
Db 21 LysAlaAsnGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMe 40
QY 2005 GCAAACTTCAAAG---CAGCCCTCCGAGAGCATTGCTTCGAGCCCTCCATGAAA-- 2059
Db 40 tGlnThrPheLysAlaGluProGluLysProSerAlaPheGluProAlaIleGluMe 60
QY 2060 -TCAAAAGTCTGCCAAATAAAC---CTTGGAATTAAGATGAACAACATTGAGGCAGA 2115
Db 60 tGlnLysSerValProAsn-LysAlaLeuGluLysAsnGluInThrLeuArgAla 80
QY 2116 TGAGAA---CTCCCATCA---AATCCAAACAAAGAGACTATGAGAAAGTCTCTGGGATTCT 2169
Db 80 sp-GluIleLeuProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer 99
QY 2170 GGAGTCTCTG---TAGACTGTTTCCAGAGAGGTGGT---TTTACCAGGCTACATCA 2223
Db 100 -GluSerLeuGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisG1 119
QY 2224 AAAAGTAATGATGATAA---TAATGGAATAATGAGAGTCTCT-CATAATGAGTGTCTG 2279
Db 119 nLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLe 139
QY 2280 AAGCTCCCT---CAGAAATGAAAGTCTATTCCAACTAAAGCCT---TGAATGATGGCAT 2333
Db 139 uLysAlaProCysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMe 159
QY 2334 GCAAACTTCAAAGCAGCCTCCG---AGAGCATCTGCTTCGAGCCTGCATTGAAAT- 2389
Db 159 tGlnThrPheLysAlaGluProGluLysProSerAlaPheGluProAlaIle-GluM 179
QY 2390 --GAAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAGATGAACAACATTGAGCAGA 2444
Db 179 etGlnLysSerValProAsn-LysAlaLeuGluLysAsnGluInThr---LeuArg 197

```

RESULT 13

US-09-433-826B-473

; Sequence 473, Application US/09433826B

; Patent No. 6579973

; GENERAL INFORMATION:

; APPLICANT: Jiang, Yuqui

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.470C4

; CURRENT APPLICATION NUMBER: US/09/433,826B

; CURRENT FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 474

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 473

; LENGTH: 445

; TYPE: PRN

; ORGANISM: Homo sapiens

```
US-09-433-826B-473
Alignment Scores:
Pred. No.: 3,72e-38 Length: 445
Score: 421.50 Matches: 226
Percent Similarity: 61.49% Conservative: 55
Best Local Similarity: 49.45% Mismatches: 119
Query Match: 5.95% Indels: 60
DB: 4 Gaps: 38

US-09-602-362E-22 (1-4115) x US-09-433-826B-473 (1-445)
QY 1894 AAAAGATAGATAAAATAAGTAAAGGGT---CCCTGTTAAAGTGTCTCTG 1950
D 1 LysGluIleAspLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeu 20
QY 1951 AGGCTAACT---GGGAATGAAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACAT 2004
D 21 LysAlaAsnCysGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMe 40
QY 2005 GCAAACTTCAAG---CAGGCTCCGAGAGCCATCTGCTTCGAGCTCCATTGAAA--- 2059
D 40 tGlnThrPhelYsAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMe 60
QY 2060 -TCAAAAGTCTGCTCAATAAAC---CTTGGAATTAAGATGAACAAACATTGAGGCGAGA 2115
D 60 tGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaA 80
QY 2116 TGAGAA---CTCCCATCA---AATCCAAACAAGGACTATGAGAAAGTTCTTCGGATTCT 2169
D 80 sp-GluIleLeuProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer 99
QY 2170 GGAGTCTCTG---TAGACTGTTTCCAGAGAGTGGT---TTTACCAGGCTACACATCA 2223
D 100 -GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisG 119
QY 2224 AAAAGATAGATAAAA---TAATGAAATAGAGAGTCTCT-GATAATGAGGTTTCTG 2279
D 119 nLysGluIleAspLysIleAsnGlyLysLeuGluGlySerProAspAsnAspGlyPheLe 139
QY 2280 AAGCTCCCT---GAGATGAAAGTTCTATCCACTAAAGCCT---TGAATGATGGCAT 2333
D 139 uLysAlaProCysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMe 159
QY 2334 GCAAACTTTCAAGACAGGCTCCCG---AGAGCCATCTGCTCGAGCTGCTCATTTGAAAT- 2389
D 159 tGlnThrPhelYsAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluM 179
QY 2390 -GAAAGTCTGTTCCAAATAAC---CTTGGAATTAAGATGAACAAACATTGAGACAGA 2444
D 179 etGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArg 197
QY 2445 TCAGAT-----TTCCTCTCAGATCAAAACAAGAGGTTGAGAAA---TTCTGGGAT 2495
D 198 AlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsnSerTrpAsp 217
QY 2496 TCTGAAGTCT---CCGTGGAGTCTTTCCAGAGAGGATGGTGTATCC---CAGCTACAC 2549
D 218 SerGlu-SerLeuArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrH 237
QY 2550 ACAAAGAAAGGATAAAT---AGTGGAAATTAAGAGATTTCACATAGCTATCAAAAT- 2605
D 237 isGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysI 257
QY 2606 --CTTGTACAGTTTCAT---CTTGTAAGAGCGAAGGAACTTCAAAAAGACACTGTGAAC 2660
D 257 leLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluG 277
QY 2661 A-CGTACAGGAAATGGAACAAT---GAAAGAGATTTTGTGTACTGAAAAAGAAC---T 2713
D 277 lnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuS 297
QY 2714 GTCAGACAAAGAAATAATATCACAGTAGAGAA---CCAAAAGTTAAATGGGAACAAGAG 2770
```

```

297 exGluAlaLysGluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGlu 316
2771 ---TCTGCAGTGTGATGACTTAAACCAA---GAAAGAGAGAGAGAGAAATGCCGATAT 2824
317 LeuCysSerValArgLeuThrLeuAsnGlnGluGluGluLysArg-ArgAsnAlaAspI 336
2825 ATTAAGAAAAA---TTAGGAAGAATTAGAGAATCCGAGACGACATAGA 2872
336 eLeuAsnGluLysIleArgGluLeuGlyArgIle-GluGluGlnHisArg-----L 354
2873 AAGAGTTAAGAGTGAACACAACTTGAAG---GCTCTCAGATCAACAAG---ATAAGAATT 2926
354 ysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeu 373
2927 GAAGGTGTAGAAAGTAATTTGAATAGGTTTCTCA---ACTCATCAAAATGAAA---ATTAT 2980
374 LysSerValGluSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLe 393
2981 TCTTACATCAAAATTCATGCTTGAAGAGGAATTGCCA---TGCAAACTGGAATAGCCAC 3037
393 uLeuHisGluAsnCysMetLeuLysLysGluIleAlaValMetLeuLysLeuGluIleAla 413
3038 ACTA---AACACCAATCCAGGAAAGGAAAAATAATC-TTTGAGGACTTAAGATTTTAAA 3093
413 rLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuL 433
3094 GA---AAGATGCTGAACCTAGA 3113
433 ysGluLysAsnAlaGluLeuGln 440

RESULT 14
US-09-604-287A-473
Sequence 473 Application US/09604287A
Patent No. 6586572
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
APPLICANT: Harlocker, Susan L.
APPLICANT: Hepler, William F.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.470C7
CURRENT APPLICATION NUMBER: US/09/604,287A
CURRENT FILING DATE: 2000-06-22
NUMBER OF SEQ ID NOS: 489
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 473
LENGTH: 445
TYPE: PRT
ORGANISM: Homo sapiens
US-09-604-287A-473

Alignment Scores:
Pred. No.: 3,72e-38 Length: 445
Score: 421.50 Matches: 226
Percent Similarity: 61.49% Conservative: 55
Best Local Similarity: 49.45% Mismatches: 119
Query Match: 5.95% Indels: 60
DB: 4 Gaps: 38

US-09-602-362E-22 (1-4115) x US-09-604-287A-473 (1-445)
QY 1894 AAAAGATAGATAAAATAAGTAAAGGGT---CCCTGTTAAAGTGTCTCTG 1950
D 1 LysGluIleAspLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeu 20
QY 1951 AGGCTAACT---GGGAATGAAAGTTCTATTCCAACTAAAGCCTAGA---ATTGATGACAT 2004
D 21 LysAlaAsnCysGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMe 40
QY 2005 GCAAACTTCAAG---CAGGCTCCGAGAGCCATCTGCTTCGAGCTCCATTGAAA--- 2059
D 40 tGlnThrPhelYsAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMe 60
QY 2060 -TCAAAAGTCTGCTCAATAAAC---CTTGGAATTAAGATGAACAAACATTGAGGCGAGA 2115
D 60 tGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaA 80
QY 2116 TGAGAA---CTCCCATCA---AATCCAAACAAGGACTATGAGAAAGTTCTTCGGATTCT 2169
D 80 sp-GluIleLeuProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer 99
QY 2170 GGAGTCTCTG---TAGACTGTTTCCAGAGAGTGGT---TTTACCAGGCTACACATCA 2223
D 100 -GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisG 119
QY 2224 AAAAGATAGATAAAA---TAATGAAATAGAGAGTCTCT-GATAATGAGGTTTCTG 2279
D 119 nLysGluIleAspLysIleAsnGlyLysLeuGluGlySerProAspAsnAspGlyPheLe 139
QY 2280 AAGCTCCCT---GAGATGAAAGTTCTATCCACTAAAGCCT---TGAATGATGGCAT 2333
D 139 uLysAlaProCysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMe 159
QY 2334 GCAAACTTTCAAGACAGGCTCCCG---AGAGCCATCTGCTCGAGCTGCTCATTTGAAAT- 2389
D 159 tGlnThrPhelYsAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluM 179
QY 2390 -GAAAGTCTGTTCCAAATAAC---CTTGGAATTAAGATGAACAAACATTGAGACAGA 2444
D 179 etGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArg 197
QY 2445 TCAGAT-----TTCCTCTCAGATCAAAACAAGAGGTTGAGAAA---TTCTGGGAT 2495
D 198 AlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsnSerTrpAsp 217
QY 2496 TCTGAAGTCT---CCGTGGAGTCTTTCCAGAGAGGATGGTGTATCC---CAGCTACAC 2549
D 218 SerGlu-SerLeuArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrH 237
QY 2550 ACAAAGAAAGGATAAAT---AGTGGAAATTAAGAGATTTCACATAGCTATCAAAAT- 2605
D 237 isGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysI 257
QY 2606 --CTTGTACAGTTTCAT---CTTGTAAGAGCGAAGGAACTTCAAAAAGACACTGTGAAC 2660
D 257 leLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluG 277
QY 2661 A-CGTACAGGAAATGGAACAAT---GAAAGAGATTTTGTGTACTGAAAAAGAAC---T 2713
D 277 lnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLysLysLeuS 297
QY 2714 GTCAGACAAAGAAATAATATCACAGTAGAGAA---CCAAAAGTTAAATGGGAACAAGAG 2770
```

Qy	2005	GCAACCTTCAAAG---CAGGCTCCCGAGAGCCACTGCTCCCTTCAGAGCTCCATTTGAAA--	2059
Db	40	tGlnThrPhenylsAlaGluProProGluLysProSerAlaPheGluProAlaIleGluLeu60	
Qy	2060	-----TCAAAAGTCTGCCAATAAAC---CTTGGAAATTAACAATGAACAACAATTTGAGGCAGA2115	
Db	60	tGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuAlaAla80	
Qy	2116	-----TGAGAA---CTTCCCATCA---AATCCAAACAAGAGCATATGAGAAAGTTCTTGGGATTTCT2169	
Db	80	sp-GluIleLeuProSerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSer99	
Qy	2170	GGAGTCTCTG---TAGACTGTTTCCAGNAGATGGTG---TTTACCAGAGCTACACATCA2223	
Db	100	-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisGln119	
Qy	2224	AAAAAGAAATAGATAAAA---TAATGGAAAAATAGAGAGTCTCT-CATTAATGAGGTTTCTG2279	
Db	119	nLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLe133	
Qy	2280	AAGGCTCCCT---CAGAAATGAAAGTTCTATTCCACTAAAGCT---TGAATTGATGGCAT2333	
Db	139	uLysAlaProCysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMe159	
Qy	2334	GCAAACTTTCAAAGCAGCGCTCCCG---AGAGGCCATCTGCTTCAGAGCTGCATGAAT-2389	
Db	159	tGlnThrPhenylsAlaGluProProGluLysProSerAlaPheGluProAlaIle-GluM179	
Qy	2390	---GAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAGATGAACAACATTTGAGACAGA2444	
Db	179	etGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThr---LeuArg197	
Qy	2445	TCAGAT-----TTCCCTTCAGAACTCAAAACAAGAAGGTTTGAAAA---TTCTGGGAT2495	
Db	198	AlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsnSerTrpAsp217	
Qy	2496	TTGGAAGTCT---CCGTGGAGTGTTCACAGAAGGATGGTGTGTACC---CAGGCTACAC2549	
Db	218	SerGlu-SerLeuArgGluThrValSerGlnLysAsp-ValCysValProLysAlaThrHr237	
Qy	2550	ACAAAAGAAAGGATAAAAT---AGTGGAAAAATTAGAGATTCCTACTAGCCTATCAAAAT-2605	
Db	237	isGlnLysGluMetAspLysIleSerGlyLysLeuGluLysAspSerThrSerLeuSerLysI257	
Qy	2606	--CTTGTACAGTTTCAT---CTTGTGAAAGCGCAAGGAACTTCAAAAAGACACTGTGCAAC2660	
Db	257	leLeuAspThrValHisSerCysGluAlaAlaArgGluLeuGlnLysAspHisCysGluG277	
Qy	2661	A-CGTACAGAAAAATGGAACAAT---GAAAAAGAGTTTGTGTACTGAAAAAGAAC--T2713	
Db	277	lnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLysLeuS297	
Qy	2714	GTCAACAACAAAAGAAATAAATTCAGTAPAGAAA---CCAAAGTTAAATCGGAACAAGAG2770	
Db	297	erGluAlaLysGluIleLysSerGln-LeuGluAsnGlnLysValLysTrpGluGlnGlu316	
Qy	2771	---TCTGCAGTGTAGATTGACTTAACCAA---GAAAAGAGAGAGAGAAATCCCGATAT2824	
Db	317	LeuCysSerValArgLeuThrLeuAsnGlnGluGluLulLysArg-ArgAsnAlaAspI336	
Qy	2825	ATTAAGAAAAA---TTAGAGCAAAATTAGAGATTCGAGAGCAGCATAGA2872	
Db	336	eLeuAsnGluLysIleArgGluLeuGluGlyArgIle-GluGluGlnHisArg-----L354	
Qy	2873	AAGAGTTAGAGTGAAACACAACTTGAAG---GCTCTCAGATACAG---ATAAGAAAT2926	
Db	354	ysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeuArgIleGlnAspIleGluLeu373	
Qy	2927	GAAGGTGTACAAAGTAATTTGAATAGTTTCTCA---ACTCATGAATGAAA---ATTAT2980	
Db	374	LysSerValGluSerAsnLeuAsnGln-ValSerHisThrHisGluAsnGluAsnTyrLe393	
Qy	2981	TCATTACATGAAATTCGATGTTGAAAAAGGAATTTGCCA---TGCAAAACTCGAATAGCCAC3037	

```

Db      393 uLeuHisGluAsnCysMetLeuLysGlyGluIleAlaMetLeuLysLeuGluIleAlaTh 413
      :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      3038 ACTA---AACACCAATCAGGAAAGGAAATAAATC-TTTGAGCACTTAAGATTTTAAA 3093
      ||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      413 rLeuLysHis-GlnTy-rGlnGluLysGluAsnLysTy-rPheGluAspIleLysIleLeuL 433
      :||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      3094 GA---AAGATGCTGAACCTTAGA 3113
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      433 ySGluLysAsnAlaGluLeuGln 440
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 15
US-09-834-759-473
; Sequence 473, Application US/09834759
; Patent No. 6880197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William F.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 473
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-759-473

Alignment Scores:
Pred. No.: 3,72e-38 Length: 445
Score: 421.50 Matches: 226
Percent Similarity: 61.49% Conservative: 55
Best Local Similarity: 49.45% Mismatches: 119
Query Match: 5.95% Indels: 60
DB: 4 Gaps: 38

US-09-602-362E-22 (1-4115) x US-09-834-759-473 (1-445)
QY      1894 AAAAGATAGATATAATAATGGAATAATGAAAGGT---CCCTGTTAAAGTCGTCTCTG 1950
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1 LysGluIleAspLysIleAsnGlyLysLeuGluGlySerProValLysAspGlyLeuLeu 20
      :||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      1951 AGGCTAACT---GGGAATGAAAGTTCATTCCAACTAAAGCCCTAGA---ATTGATGACAT 2004
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      21 LysAlaAsnCysGlyMetLysValSerIleProThrLysAla-LeuGluLeuMetAspMe 40
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      2005 GCAAACTTCAAAG---CAGCCCTCCGAGAGCACTCGCTCGAGCCCTCCATTGAAA-- 2059
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      40 tGlnThrPheLysAlaGluProGluLysProSerAlaPheGluProAlaIleGluMe 60
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      2060 -TCAAAAGTCGTCCAAATAAAC---CTTGGAAATTAAAGATGAACAAACATTGAGGACGA 2115
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      60 tGlnLysSerValProAsn-LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAla 80
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      2116 TGAGAA---CTCCCATCA---AATCCAAACAAGGACTATGAGAAAGTCTTGGGATTC 2169
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      80 sp-GluIleLeuProSerGluSerLysGlnLysAspTy-rGluLysSerTrpAspSer 99
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      2170 GGAGTCTCTG---TAGACTGTTTCCAGAGGATGGTG---TTTACCCAGGCTACACATCA 2223
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      100 -GluSerLeuLysGluThrValSerGlnLysAspValCysLeuProLysAlaAlaHisG 119
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      2224 AAAAGATAGATAAAA---TAATGGAAATAGAGAGTCTCT-GATAATCAGGTTTCTG 2279
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      119 nLysGluIleAspLysIleAsnGlnGlyLysLeuGluGluSerProAspAsnAspGlyPheLe 139
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

Search completed: July 15, 2004, 09:28:02
Job time : 112.997 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:47:52 ; Search time 31.2381 Seconds
(without alignments)

12140.450 Million cell updates/sec

Title: US-09-602-362E-26

Perfect score: 6448

Sequence: 1 caagagcttgccgatacaga.....tgacttctaattgaaggaa 3673

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09602362/runat_15072004_093627_22040/app_query.fasta.1.10325
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLCN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09602362 @CGN_1_133 @runat_15072004_093627_22040 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:
1: /cgn2_6/prodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/prodata/2/iaa/6C COMB.pcp.*
6: /cgn2_6/prodata/2/iaa/6D COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	2372.5	36.8	1095	4	US-09-620-405B-493
2	2372.5	36.8	1095	4	US-09-620-405B-493
3	2367.3	36.7	1002	4	US-09-620-405B-475
4	2367.3	36.7	1002	4	US-09-620-405B-475
5	2367.3	36.7	1002	4	US-09-620-405B-475
6	1775	27.5	743	4	US-09-620-405B-494
7	1775	27.5	743	4	US-09-620-405B-494
8	1769.5	27.4	650	4	US-09-620-405B-469
9	1769.5	27.4	650	4	US-09-620-405B-469
10	1769.5	27.4	650	4	US-09-620-405B-469
11	1769.5	27.4	650	4	US-09-620-405B-469
12	1414.5	21.9	445	4	US-09-620-405B-473

13	1414.5	21.9	445	4	US-09-433-826B-473
14	1414.5	21.9	445	4	US-09-604-287A-473
15	1414.5	21.9	445	4	US-09-834-759-473
16	1384	21.5	466	4	US-09-620-405B-472
17	1384	21.5	466	4	US-09-433-826B-472
18	1384	21.5	466	4	US-09-604-287A-472
19	1384	21.5	466	4	US-09-834-759-472
20	884.5	13.7	432	4	US-09-389-661-181
21	884.5	13.7	432	4	US-09-620-405B-181
22	884.5	13.7	432	4	US-09-339-338-181
23	884.5	13.7	432	4	US-09-433-826B-181
24	884.5	13.7	432	4	US-09-604-287A-181
25	884.5	13.7	432	4	US-09-834-759-181
26	798.5	12.4	228	4	US-09-620-405B-470
27	798.5	12.4	228	4	US-09-433-826B-470
28	798.5	12.4	228	4	US-09-604-287A-470
29	798.5	12.4	228	4	US-09-834-759-470
30	798.5	12.4	228	4	US-09-439-313-378
31	700.5	10.9	1719	4	US-09-352-616A-378
32	700.5	10.9	1719	4	US-09-636-215-378
33	700.5	10.9	1719	4	US-09-685-166A-378
34	586.5	9.1	671	4	US-09-439-313-380
35	586.5	9.1	671	4	US-09-352-616A-380
36	586.5	9.1	671	4	US-09-289-198-306
37	586.5	9.1	671	4	US-09-636-215-380
38	586.5	9.1	671	4	US-09-685-166A-380
39	586.5	9.1	671	4	US-09-439-313-379
40	586.5	9.0	656	4	US-09-636-215-379
41	583.5	9.0	656	4	US-09-352-616A-379
42	583.5	9.0	656	4	US-09-289-198-305
43	583.5	9.0	656	4	US-09-636-215-379
44	583.5	9.0	656	4	US-09-685-166A-379
45	583.5	9.0	656	4	US-09-685-166A-379

ALIGNMENTS

RESULT 1

US-09-620-405B-493
; Sequence 493, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620.405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(1095)
US-09-620-405B-493

Alignment Scores:	6.47e-183	Length:	1095
Pred. No.:	2372.50	Matches:	529
Score:	59.3%	Conservative:	73
Percent Similarity:	52.17%	Mismatches:	151
Best Local Similarity:	36.79%	Indels:	261
Query Match:	4	Gaps:	13
DB:			

Qy	2960	CCAATAAAGCCTTAGAATGGAGCAATAACAACATTTGAGAGCA-----	3004
Db	589	ProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePro	608
Qy	3004	-----	3004
Db	609	SerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGluSerLeuArgGlu	628
Qy	3004	-----	3004
Db	629	ThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys	648
Qy	3005	-----GATTCAACTACCCATTCAAAATCTTGGATGCACATTCCTCTCT	3046
Db	649	IleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHisSer	668
Qy	3047	TGTGAAGAGGAGGGGAACCTTAAAAAGATAAAGCTGTGAACAAATTCAGACAAAATGGAA	3106
Db	669	CysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMetGlu	688
Qy	3107	CAAAATGAAAAATAGTTTCTGTACTACAAAGGAAGAACTGTGCAGAGCGAAGAATAAAA	3166
Db	689	GlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluIleLys	708
Qy	3167	TCACAGTTAGAGAACCCAAAAGCTTAAATGGGAAACAAGAGCTCTGCAGTGTGAGATTCCCT	3226
Db	709	SerGlnLeuGluAsnGlnLysValLysTrpGluGlnLeuCysSerValArgLeuThr	728
Qy	3227	TTAAATCAAGAAGAGAGAGAGAAATGTCGATATATTAAAGAAAAAATTAGACCC	3286
Db	729	LeuAsnGlnGluGluGluLysArgArgAsnAlaAspIleLeuAsnGluLysIleArgGlu	748
Qy	3287	-----GAAGAGCACTTTAGGAAAAAGTTTAGAAGTGAACACCAACTTGAA	3331
Db	749	GluLeuGlyArgIleGluGluGlnHisArgLysGluLeuGluValLysGlnLeuGlu	768
Qy	3332	CAGACTCTCAGAATACAAGATATAGAATTGAAAGTGTACAAAGTAATTTGAATCAGGTT	3391
Db	769	GlnAlaLeuArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGlnVal	788
Qy	3392	TCTCACACTCATGAAGTGAATAATGATCTCTTTTCATGAAAATTCGATGTTTGAAAAGGAA	3451
Db	789	SerHisThrHisGluAsnGluAsnTrpLeuLeuHisGluAsnCysMetLeuLysLysGlu	808
Qy	3452	ATTGCCATGCTAAACCTGGAAGTAGCCCACTGAAACATCAACACCAAGGTGAAGGAAAAT	3511
Db	809	IleAlaMetLeuLysLeuGluIleAlaThrLeuLysHisGlnTrpGlnGluLysGluAsn	828
Qy	3512	AAATACCTTTGAGGACATTAGATTTTACAGAAAAGAAATCCTGAACTTCAATGACCCCTA	3571
Db	829	LysTyrPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeu	848
Qy	3572	AAACTGAAACAGAAAAACAGTAAACAAAAGGGCATCTCAGTATATAGAGACGAGCTTAAAGTT	3631
Db	849	LysLeuLysGluGluSerLeuThrLysArgAlaSerGlnTrpSerGlyGlnLeuLysVal	868
Qy	3632	CTGACGGCAGAGAACACGATCTGACTTCTAAATTTGAAGGAA	3673
Db	869	LeuIleAlaGluAsnThrMetLeuThrSerLysLeuLysGlu	882

RESULT 2

	US-09-834-759-493	
	; Sequence 493, Application US/09834759	
	; Patent No. 6680197	
	; GENERAL INFORMATION:	
	; APPLICANT: Jiang, Yuqiu	
	; APPLICANT: Dillon, Davin C.	
	; APPLICANT: Mitcham, Jennifer L.	
	; APPLICANT: Xu, Jiaangchun	
	; APPLICANT: Harlocker, Susan L.	
	; APPLICANT: Hepler, William T.	
	; APPLICANT: Henderson, Robert A.	
	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND	

1622 TCCAAACGAGGAGGAGATGAAGAAATATCTTTGGGATTTCTGGGAGTCTCTTTGAGATTCT 1681
Db SerLysGlnGluGluAspGluGluTyrSerCysAspSerArgSerLeuPheGluSerSer 189
1682 GCAAGACTCAAGTGTCTATACCTGAGTCTATGATCAGAAAGTAAATGGAGATAAATAGA 1741
Db AlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnArg 209
1742 GAAGTAGAAGAGCTTCTCGAGAGCCATCTGCCTTCAAGCTCCCGTNGAAATGCAAAAG 1801
Qy GluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsn 229
1802 ACTGTTCAAATAAGCCTTTGAATTGAAGATGAACAACATTCAGAGCAGCTCAGATG 1861
Db SerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspProMet 249
1862 TTCCCATCAGAAATCCAAACAAAGGAGCGATCAAGAAATTTCTTGGGATTTCTGAGAGTCCC 1921
Db PheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSerLeu 269
1922 TGTGACAGGTTTCAAGAGGATGTGTATTTACCCAAAGCTACACATCAAAAGAAATTC 1981
Db CysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIle 289
1982 GATACCTTAAGTGAAATAGAGAGCTCTCTGTTAAAGATGCTCTTCTGAAGCCTACC 2041
Qy AspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThr 309
2042 TGTGGAAGGAAGTTTCTTTCCAAATAAGCCCTTAGAATTAAGGACAGAGAAACATTC 2101
Db CysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 329
2102 AAAGCAGAGTCTCTGATAAGATGGTCTTCTGAGCCTACCTGCTGGAAGAAAGTTCT 2161
Qy LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
2162 CTTCCAAATAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGCAGAGTCTCTCGAT 2221
Db ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp----- 366
2222 AATGATGCTCTCTGAGCCTACCTGCTGGAAGGAAAGTTCTCTTCCAAATAAGCTTTA 2281
Qy 366 ----- 366
2282 GAATTGAAGGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAATCCAAACA 2341
Db 367 -----GluIleLeuProSerGluSerLysGln 375
2342 AAGGATGATGAAGAAATCTTTGGGATTTTGGAGTTTCTTGGAGTCTCTTACAGAT 2401
Qy 376 LysAspTyrGluGluSerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLys 395
2402 GATGTGCTTTACCAAGCTACACATCAAAAGAAATTCGATACCTTAAGTGGAGAAATTA 2461
Db AspValCysLeuProLysAla**HisGlnLysGluLeuAspLysIleAsnGlyLysLeu 415
2462 GAAGAGTCTCTGATAAGATGGTCTTCTGAGCCTACCTGGAATGAAATTTCTCTT 2521
Qy 416 GluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIle 435
2522 CCAATAAGCCTTAGAATTCAGGACAGAGAAACATTCAGAGCAGAGGATGTGAGTTCT 2581
Qy 436 ProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLys 455
2582 GTAGAGTCCATTCAGTCTTTTGGCAAAACCGACTACTGAAATTCAGATCTTACAAA 2641
Qy 456 ---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal--- 469
2642 GTTGGAGAGACTTTAATCTTACTACCAGGAGGAGGACACAAAGACAGATTAACGTGGACA 2701
Db 470 -----ProAsnLysAlaLeuGluLeuLys 477

2702 CAGGAACGTGATATTGGCATTATTGAACGAGCTCCACAAGATCAACAAATAAGATGCC 2761
Qy 478 AsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 489
2762 ACATCAGAAATTTAGGAAGAAAGATACAAATACTCAACTTCAGATTCTGAGATTATCTCT 2821
Qy 490 ---SerGluSerLysGlnLysAspTyrGluGluSerTrpAspSerGluSerLeuCys 508
Db 2822 GTGACTGATACACAGATTATGAGTGTTTACCTCAGGCTACATATCAAAAGAAATAAAG 2881
Qy 509 GluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIleAsp 528
Db 2882 ACAACAAATGGCAAAATAGAAAGCTT----- 2908
Qy 529 LysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCys 548
Db 2908 ----- 2908
Qy 549 ArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLys 568
Db 2909 -----CCTGAAAGCCTTCTCACCTTCAGGCTGCGCACTGAAATGCAAAATCTGTT 2959
Qy 569 AlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 588
Db 2960 CCAATAAAGGCTTAGAATGGAGAAATAACAACATTCAGAGCA----- 3004
Qy 589 ProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePro 608
Db 3004 ----- 3004
Qy 609 SerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGluSerLeuArgGlu 628
Db 3004 ----- 3004
Qy 629 ThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys 648
Qy 3005 -----GATTCAACTACCTATCAAAATCTTGGATGCACTTCTCTCT 3046
Db 649 IleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHisSer 668
Qy 3047 TGTGAAAGGAGGAGAACTTAAAGAAATAACTCTGAACAAATTTACAGCAAAATGGA 3106
Db 669 CysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMetGlu 688
Qy 3107 CAAATGAAATAAGTTTGTCTACTACAAAGGAACCTGTCAGAGCCGAAAGAAATAAAA 3166
Qy 689 GlnMetLysLysPheCysValLeuLysLysLeuSerGluAlaLysGluIleLys 708
Qy 3167 TCACAGTTAGAGAACCAAAAGCTAAATGGGAACAGAGCTCTGCAGTGTGAGATTGCT 3226
Db 709 SerGlnLeuGluAsnGlnLysValLysTrpGluGlnLeuCysSerValArgLeuThr 728
Qy 3227 TTAATCAAGAGAGAGAGAGAAATGCGATATATTAAAGAGAAATTTAGACCC 3286
Db 729 LeuAsnGlnGluGluLysArgArgAsnAlaAspIleLeuAsnGluLysIleArgGlu 748
Qy 3287 -----GAAGAGCAACTTAGGAAAAAGTTAGAAAGTGAAGTGAACACCACTTCAA 3331
Db 749 GluLeuGlyArgIleGluGluGlnHisArgLysGluLeuGluValLysGlnLeuGlu 768
Qy 3332 CAGACTCTCAGAAATCAAGATATAGAAATGAAAGTGTAAACAATTTGAATCAGGTT 3391
Db 769 GlnAlaLeuArgIleGlnAspIleGluLysSerValGluSerAsnLeuAsnGlnVal 788
Qy 3392 TCTCAGCTCAGTAAGTGAATGATCTCTTTCATGAAATTTGATCTTCAAAAGAA 3451
Db 789 SerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGlu 808
Qy 3452 ATTGCGATCTGTAAGTACCCACTGAAACATCAACACAGGTGAGGAAAT 3511
Db 809 IleAlaMetLeuLysLeuGluIleAlaThrLeuLysHisGlnTyrGlnGluLysGluAsn 828
Qy 3512 AAATACTTTGAGGACATTAAAGATTTTACAAGAAAGAAATGCTGTAACCTTCAATGACCCTA 3571

Db 829 LysTyrPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeu 848
Qy 3572 AAACCTGAACAGAAACAGTAACAAAAGGCGATCTCAGTATAGACAGCAGCTTAAAGTT 3631
Db 849 LysLeuLysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGlyGlnLeuLysVal 868
Qy 3632 CTGACGCGACAGAACACAGATGCTGACTTCTAAATTAAGGAA 3673
Db 869 LeuIleAlaGluAsnThrMetLeuThrSerLysLeuLysGlu 882

RESULT 3

US-09-620-405B-475
; Sequence 475, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-620-405B-475

Alignment Scores:

Pred. No.: 1,74e-182 Length: 1002
Score: 2367.00 Matches: 511
Percent Similarity: 62.62% Conservative: 62
Best Local Similarity: 55.85% Mismatches: 126
Query Match: 36.71% Indels: 216
DB: 4 Gaps: 9

US-09-602-362E-26 (1-3673) x US-09-620-405B-475 (1-1002)

Qy 1199 TTGAGGCTACAAAGAAACATCTGAAATTTTCATGGCCAGCAAAAGAAAGATCTAGG 1258
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 20
Qy 1259 AAGATCATCTGGAGGAAAGAAACATCTGTAAGACTGAATCGCTGGCAGAGTAACA 1318
Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
Qy 1319 CCTAATAAACTGAAGTTTGGAAAAGGAACATCTAATAATGATTGATGCTCTCAAAA 1378
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
Qy 1379 GAAACATCTCAAAAAGCAAGTACAAATGTGGATGTGATGTTCTGTAGAGCCTATATTCA 1438
Db 61 GluSerSerThrLysAlaSerAlaAsn----- 69
Qy 1439 CTTTGTGGCACCGGACTATTGAAATTCACAGTGTACAAAGTTGAGGAAGACTTTAAT 1498
Db 69 ----- 69
Qy 1499 CTTGCTACCAAGATTATCTCTAAGAGTGTGACAGAAATTATACGTTTACCTGATGCT 1558
Db 69 ----- 69

Qy 1559 ACATATCAAAAAGATATCAAAAACAATAAATCACAAAATAGAAGATCAGATGTTCCTCATCA 1618
Db 70 -----AspGlnArgPheProSer 75
Qy 1619 GAATCCAAACGAGAGAAAGATGAAGAATATTCTTGGGATTTCTGGAGTCTCTTTGAGAGT 1678
Db 76 GluSerLysGlnGluAspGluLysSerCysAspSerArgSerLeuPheGluSer 95
Qy 1679 TCTGCAAAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAAGTATCGAGATAAAT 1738
Db 96 SerAlaLysIleGlnValCysIleProGluSerIleThrGlnLysValMetGluIleAsn 115
Qy 1739 ACAGAAGTAGAAGACGCTTCTCGAGAAGCCATCTGCCTTCAAGCCCTGCGCTNGAAATGCAA 1798
Db 116 ArgGluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln 135
Qy 1799 AAGACTGTTCCAAATAAAGCCCTTTGAATTAAGAATTAAGAAACAACTTTCAGAGCAGCTCAG 1858
Db 136 AsnSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspPro 155
Qy 1859 ATGTTTCCCATCAGAATCCAAACAAAAGGACGATGAAGAAATTTCTTGGGATTTGAGAGT 1918
Db 156 MetPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSer 175
Qy 1919 CCTGTGAGACGCTTTCACAGAAGATGTGTATTTACCCAAAGCTACACATCAAAAAGAA 1978
Db 176 LeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu 195
Qy 1979 TTCTGATACCTTAAGTGGAATTTAGACAGTCTCTGTTTAAAGATGCTTCTCTGAGGCT 2038
Db 196 IleAspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAla 215
Qy 2039 ACCTGTGGAAGGAAAAGTTTCTTCTCCAAATAAAGCCTTTAGAAATTAAGAGCAGAGAAACA 2098
Db 216 ThrCysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThr 235
Qy 2099 TTCAAAGCAGAGTCTCTGTATAAAGATCGTCTTCTGAAGCCCTACCTGTGGAGGAAGATT 2158
Db 236 PheLysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLys 255
Qy 2159 TCTTCTTCCAAATAAAGCCTTAGAATTAAGAGCAGAGAAACACTCAAAAGCAGAGTCTCCT 2218
Db 256 SerValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp----- 273
Qy 2219 GATAATGATGTTCTTCTGAAGCCTACCTGTGGAGGAAGATTCTCTCCAAATAAAGCT 2278
Db 273 ----- 273
Qy 2279 TTAGAATTGAGGACAGAGAAACATTCAAAGCAGCTCAGATGTTTCCCATCAGAAATCCAAA 2338
Db 274 -----GluIleLeuProSerGluSerLys 281
Qy 2339 CAAAAGGATGATGAAGAAAATTTTGGGATTTTGGAGTTTCTTCTTGAGACTCTCTTAAG 2398
Db 282 GlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeuCysGluThrValSerGln 301
Qy 2399 AATGATGTGTGTTTACCCAGGCTACACATCAAAAGAAATTCGATACCTTAAGTGGAAA 2458
Db 302 LysAspValCysLeuProLysAla***HisGlnLysGluIleAspLysIleAsnGlyLys 321
Qy 2459 TTAGAAGAGTCTCTCTGATAAAGATGCTTCTTGAAGCCCTACCTGTGGAAATGAAATTTCT 2518
Db 322 LeuGluLysSerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSer 341
Qy 2519 CTTTCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATTCAAAGCAGAGAGATGTGAGT 2578
Db 342 IleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGlu 361
Qy 2579 TCTGTAGAGTCCACATTCAGTCTTTTGGCAAAACCGACTACTCTGAAATTCACAGTCTACA 2638
Db 362 Lys---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal 376
Qy 2639 AAAGTTGAGGAGACTTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACTGGA 2698

```
Db 377 - - - - -ProAsnLysAlaLeuGluLeu 383
Qy 2699 CAACAGAACGTGATATTGGCATTATTGAACGAGCTCCACAAGATCAAAATAAGATG 2758
Db 384 LysAsnGluGlnThrLeuArgAlaAspGluLeuPro - - - - - 396
Qy 2759 CCCACATCAGAAATAGGAAGAAAGAGATACAAAATCAACTTCAGATTCTGAGATTATC 2818
Db 397 - - - - -SerGluSerLysGlnLysAspLysGluGluSerSerTrpAspSerGluSerLeu 414
Qy 2819 TCTGTGAGTGATACACAGAATTATGAGTGTGTTACTGAGCTACATATCAAAAAGAAATA 2878
Db 415 CysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluLe 434
Qy 2879 AAGACAATGCAAAATAGAAGTCT - - - - - 2908
Db 435 AspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaPro 454
Qy 2908 - - - - - 2908
Db 455 CysArgMetLysValSerIleProThrLysAlaLeuLeuMetAspMetGlnThrPhe 474
Qy 2909 - - - - -CTGAAAAGCTTCTCACTTGTAGCCTGCCACTGGAATGCAAACTCT 2956
Db 475 LysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSer 494
Qy 2957 GTTCCAAATAAGCTTAGAATGGAAGATAAAACAAACATTGAGAGCA - - - - - 3004
Db 495 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPhe 514
Qy 3004 - - - - - 3004
Db 515 ProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGluSerLeuArg 534
Qy 3004 - - - - - 3004
Db 535 GluThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAsp 554
Qy 3005 - - - - -GATTCAACTACCTATCAAAATCTTGGATGCACTTCT 3043
Db 555 LysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHis 574
Qy 3044 TCTTGTGAAGAGGAGGAACCTTAAAGAAATCACTGTGAACAAATTAACAGCAAAATG 3103
Db 575 SerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMet 594
Qy 3104 GAACAAATGAAAATAAGTTTGTGTACTACAAAGGAACCTGTGAGAGCAAGCAAGAAATA 3163
Db 595 GluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluLe 614
Qy 3164 AAATCAGATTAGAGACCAAAAGCTAAATGGGAACAAGAGCTTGTGATGTGAGATTG 3223
Db 615 LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerValArgLeu 634
Qy 3224 CCTTTAATCAAGAGAGAGAGAGAGAAATGTCATATATTAAAGAAAATAATAGA 3283
Db 635 ThrLeuAsnGlnGluGluLysLysPheArgAsnAlaAspIleLeuAsnGluLysIleArg 654
Qy 3284 CCC - - - - -GAAGACAACTTAGGAAAAGTTAGAGTGAACACCAACTT 3328
Db 655 GluGluLeuGlyArgIleGluGluGlnHisArgLysGluLeuGluValLysGlnGlnLeu 674
Qy 3329 GAACAGACTCTCAGAAATACAGATAGAAATGGAAGTGTACAGTAATTTGATCAG 3388
Db 675 GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln 694
Qy 3389 GTTCTCAGACTCATGAAAGTGAATGATCTCTTTTCATGAAAATGTCATGTTGAAAAG 3448
Db 695 ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLys 714
Qy 3449 GAAATGGCATGCTAAACCTGGAAGTAGCCACACTGAAACATCAACACAGAGTGAAGAA 3508
```

```
Db 715 GluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHisGlnTyrGlnLysGlu 734
Qy 3509 RATAAATCTTTGAGGACATTAAGATTTTACAAGAAAAGAAATGCTGAACCTTCAATGACC 3568
Db 735 AsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThr 754
Qy 3569 CTAAACTGAAACACAGAAACAGTAAACAAAAGGCATCTCAGTATAGAGACAGCTTAAA 3628
Db 755 LeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGlyGlnLeuLys 774
Qy 3629 GTTCTGACGGCAGAGAACACCATGCTGCTCTTAAATTGAGGAA 3673
Db 775 ValLeuIleAlaGluAsnThrMetLeuThrSerLysLeuLysGlu 789

RESULT 4
US-09-604-287A-475
; Sequence 475, Application US/09604287A
; Patent No. 696572
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillion, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604/287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-604-287A-475

Alignment Scores:
Pred. No.: 1,748-182 Length: 1002
Score: 2367.00 Matches: 511
Percent Similarity: 62.62% Conservative: 62
Best Local Similarity: 55.85% Mismatches: 126
Query Match: 36.71% Indels: 216
DB: 4 Gaps: 9

US-09-602-362E-26 (1-3673) x US-09-604-287A-475 (1-1002)
Qy 1199 TTGAGCGCTACAAAGAAACATCTGAGAAATTTTCATGCCAGCAAAAGAGATCTAGG 1258
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrLysAlaLysGlyArgProArg 20
Qy 1259 AAGATCACATGGAGCAAAAGAAACATCTGTAAGACTGAATGCTGCCAGGAGTAAACA 1318
Db 21 LysIleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThr 40
Qy 1319 CCTATAAATCTGAAGTTTGGAAAAAGCAATCATATATGTTGCTGCTCTACAAA 1378
Db 41 SerAsnLysThrLysValLysGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
Qy 1379 GAACATCTCAAAAAGCAAGTACAAATCTGGATGTGAGTCTCTGTAGAGCCCTATATTCAGT 1438
Db 61 GluSerThrLysAlaSerAlaAsn - - - - - 69
Qy 1439 CTTTGTGGCACAGGACTATTGAAAATTCACAGTGTACAAAGTTGAGGAAGACTTTAAT 1498
Db 69 - - - - - 69
Qy 1499 CTTGCTACCAAGATTATCTTAAGAGTCTGCACAGATTATACGTGTTTACCTGATGCT 1558
```

Db 69 ----- 69
QY 1559 ACATATCAAAAGATATCAAAACAATAAATCAAAATAAGACATCAGATGTTCCATCA 1618
Db 70 -----AspGlnArgPheProSer 75
QY 1619 GAATCCAAACGAGAGAGATGAAGATATTTCTGGGATTTCTGGGAGTCTCTTTGAGAGT 1678
Db 76 GluSerLysGlnGluAspGluGluTyrSerCysAspSerArgSerLeuPheGluSer 95
QY 1679 TCTGCAAGACTCAAGTGTGTATPACTGAGTCTATGATATCAGAAAGTAAATGAGAGTAAAT 1738
Db 96 SerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsn 115
QY 1739 AGAAGAGTACAGAGCTTCTGAGAGCCATCTGAGAGCCATCAAGCCCTCCGCTNGAATGCAA 1798
Db 116 ArgGluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln 135
QY 1799 AAGACTGTTCCAAATAAAGCCTTTGAATTCAGAAATGAAACAACATTTGAGAGCAGCTCAG 1858
Db 136 AsnSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspPro 155
QY 1859 ATGTTCCCATCAAGATCCAAACAAAAGAGCAGATGAAGAAAATTTCTGGGATTTCTGAGAGT 1918
Db 156 MetPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSer 175
QY 1919 CCCTGTGAGACGGTTTCCACAGAAGGATGTATTTACCCAAAGCTACACATCAAAAAGAA 1978
Db 176 LeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu 195
QY 1979 TTCGATACCTTAAAGTGGAAATTTAGAGAGTCTCTGTTAAAGATCGTCTTCTGAGCGCT 2038
Db 196 IleAspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAla 215
QY 2039 ACCTGTGGAAGGAAGCTCTCTCCAAATAAGACCTTGAATTAAGACGACAGACAGAAACA 2098
Db 216 ThrCysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThr 235
QY 2099 TTCAAAGCAGAGTCTCTGTAAGATGCTTCTGGAAGCTACCTGTGGAAGGAAGTT 2158
Db 236 PheLysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLys 255
QY 2159 TCTCTTCCAAATAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGCAGAGTCTCT 2218
Db 256 SerValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp----- 273
QY 2219 GATAATGATGCTCTCTGGAAGCCTACCTGTGGAAGGAAGTCTCTCTTCCAAATAAGCT 2278
Db 273 ----- 273
QY 2279 TTAGAATTGAAGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAA 2338
Db 274 -----GluIleLeuProSerGluSerLys 281
QY 2339 CAAAAGGATCATGAAGAAAATTTCTGGGATTTTGGAGATTTCTTCCAGACTCTCTTACAG 2398
Db 282 GlnLysAspTyrGluGluSerTrpAspSerGluSerLeuCysGluThrValSerGln 301
QY 2399 AATGATGTGTGTTTACCCAGGCTTACACATCAAAAAGAAATTCGATACCTTAAAGTGAAAA 2458
Db 302 LysAspValCysLeuProLysAla**HisGlnLysGluIleAspLysIleAsnGlyLys 321
QY 2459 TTAGAAGAGTCTCTGATAAGATGCTTCTGGAAGCCTACCTGTGGAATGAAAATTTCT 2518
Db 322 LeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSer 341
QY 2519 CTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATTTCAAAGCAGAGATGTGAGT 2578
Db 342 IleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGlu 361
QY 2579 TCTGTAGATCCCATTCAGTCTTTTGGCAACCGAGCTACTCGAAAATTCACAGTCTACA 2638
Db 714 ----- 714

362 Lys---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal 376
2639 AAAGTTGAGGAGACTTTAATCTTACTACCAAGGAGGAGCAACAACACACAGTAACTCGA 2698
377 -----ProAsnLysAlaLeuGluLeu 383
2699 CAACAGGAACCTGATATTGGCATTATTCAACGAGCTCCACAAAGATCAACAAATAAGATG 2758
384 LysAsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 396
2759 CCCACATCAGAATTAGGAAGAAAGAGATACAAAATCAACTTCAGATTCTGAGATTATC 2818
397 -----SerGluSerLysGlnLysAspTyrGluGluSerTrpAspSerGluSerLeu 414
2819 TCTGTGAGTATACACAGAATTTATGAGTGTCTACTGAGGCTACATATCAAAAAGAAATA 2878
415 CysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIle 434
2879 AAGCAACAATAATGCAAAATAGAGTCT----- 2908
435 AspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaPro 454
2908 ----- 2908
455 CysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPhe 474
2909 -----CCTGAAAAGCCTTCTCATTCTGAGCTGCCACTGAAATAATGCAAAACTCT 2956
475 LysAlaGluProProGlyLysProSerAlaPheGluProAlaIleGluMetGlnLysSer 494
2957 GTTCCAAATAAGCCTTAGAATGGAAGATATAAACAACATTGAGAGCA----- 3004
495 ValProAsnLysAlaLeuGluLeuLysAsnGlnThrLeuArgAlaAspGlnMetPhe 514
3004 ----- 3004
515 ProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGluSerLeuArg 534
3004 ----- 3004
535 GluThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAsp 554
3005 -----GATTCAACTACCTTATCAAAAATCTTGATGCTGCACTTCTCT 3043
555 LysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHis 574
3044 TCTTGTGAAAGAGGAGGAACTTTAAAGAGATAACTGTGTGAACAAATACAGCAAAATG 3103
575 SerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMet 594
3104 GAACAAATGAAAATAAGTTTGTGTACTACAAAGGAGACTGTGAGAGCGCAAGAAATA 3163
595 GluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluIle 614
3164 AAATCACAAGTAGAGAACCAAAAAGCTAAATGGGAACAAGAGCTCTCAGTGTGAGATTG 3223
615 LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnLeuLeuCysSerValArgLeu 634
3224 CCTTTAAATCAAGAGAGAGAGAGAGAAATGTCATATATATTAAAGAAAAATTAGA 3283
635 ThrLeuAsnGlnGluGluLysArgAsnAlaAspIleLeuAsnGluLysIleArg 654
3284 CCC-----GAAGAGCAACTTAGGAAAAAGTTAGAAAGTGAAACACCAACTT 3328
655 GluGluLeuGlyArgIleGluGluGlnHisArgLysGluLeuGluValLysGlnGlnLeu 674
3329 GAACAGACTCTCAGAATAACAAGATATAGAAATGAAAAGTGTAAACAGTAATTTGAATCAG 3388
675 GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln 694
3389 GTTCTCTCAGCTCAGAAAGTCAAAATCATCTCTTTTATGAAATTCATCTGTTGAAAAG 3448
695 ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLys 714

```

QY 3449 GAAATTCGCTGCTAAACCTGGAGTAGTACCGACACTGAAACATCAACACACAGGTGAAGGAA 3508
Db 715 GluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHisGlnTyrGlnGluLysGlu 734
QY 3509 AATAAATACCTTTGAGGACATTAAGATTTTACAGAAAGAAATGCTGAACCTTCAAAATGACC 3568
Db 735 AsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThr 754
QY 3569 CTAAATCTGAACAGAAAAAGTAAACAGTAAACAAAGGCGATCTCAGTATAGAGAGAGCTTAAA 3628
Db 755 LeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGlyGlnLeuLys 774
QY 3629 GTTCTGACGCGAGACAGACGATGCTGACTCTCTAAATTTGAAGGAA 3673
Db 775 ValLeuIleAlaGluAsnThrMetLeuThrSerLysLeuLysGlu 789

RESULT 5
US-09-834-759-475
; Sequence 475, Application US/09834759
; Patent No. 5680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: DIAGNOSIS OF BREAST CANCER
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRP
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-834-759-475

Alignment Scores:
Pred. No.: 1,74e-182 Length: 1002
Score: 2367.00 Matches: 511
Percent Similarity: 62.62% Conservative: 62
Best Local Similarity: 55.85% Mismatches: 126
Query Match: 36.71% Indels: 216
DB: 4 Gaps: 9

US-09-602-362E-26 (1-3673) x US-09-834-759-475 (1-1002)
QY 1199 TTGAGGCTCAAAAGAAACATCTGAGAAATTTTCATGGCCACCAAAAGAAAGATCTAGG 1258 -
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 20
QY 1259 AAGATCATCATGGGAGGAGAAAGAAACATCTGTAAGACTGAAATGCGTCGAGAGGTAACA 1318
Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY 1319 CCTAATTAACCTGAAGTTTGGAAAAAGCAACATCTAATATGATTCGATGCTCTACAAA 1378
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
QY 1379 GAAACATCTCAAAAGCAAGTACAAATGTGGATGTGATGCTTCTGTAGAGCCTATATTCA 1438
Db 61 GluSerSerThrLysAlaSerAlaAsn 9
QY 1439 CTTTTTGGCACCGGACTATTGAAATTTACAGTGTACAAAAGTTGAGGAAGACTTTAAT 1498

```

```

Db 69 ----- 69
QY 1499 CTTGTACCAAGATTATCTCTAAGAGTGTCTGCACAGAATTATACGTGTTTACCTGATGCT 1558
Db 69 ----- 69
QY 1559 ACATATCAAAAGATATCAAAACAAATAATCACAATAAGAGATCAGATGTTCCCATCA 1618
Db 70 ----- 75
QY 1619 GAATCAAAACGAGAGAAAGATGAAGAATPATCTTTGGGATTTCTGGAGTCTCTTTGAGAGT 1678
Db 76 GluSerLysGlnGluGluAspGluLysSerCysAspSerArgSerLeuPheGluSer 95
QY 1679 TCTGCAAAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAAGTAAATCGAGATAAT 1738
Db 96 SerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsn 115
QY 1739 AGAGAAGTAGAAGAGCTTCTCTGAGAAGCCATCTGCTTCAAGCCCTGCGGTGAAATGCAA 1798
Db 116 ArgGluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln 135
QY 1799 AAGACTGTCCAAATAAAGCCTTTGAATTTGAAGAAATGAACAAACATTTGAGAGAGCTCAG 1858
Db 136 AsnSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspPro 155
QY 1859 ATGTTCCCATCAGAAATCCAAACAAAGAGCAGATCAAGAAATTTCTGGGATTTCTGAGAGT 1918
Db 156 MetPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSer 175
QY 1919 CCTGTGAGACGGTTTACAGAGAGTGTGTATTACCCAAAGCTACACATCAAAAGAA 1978
Db 176 LeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu 195
QY 1979 TTCGATACCTTAAGTGGGAAATTTAGAGAGTCTCTCTTAAAGATGCTCTTCTGAAGCCT 2038
Db 196 IleAspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAla 215
QY 2039 ACCTGTGGAAGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTTAAAGGACAGAGAAACA 2098
Db 216 ThrCysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThr 235
QY 2099 TTCAAAGCAGAGTCTCTCGATAAAGATGGTCTTCTGAGGCTACCTCTGGAAGAAAGTT 2158
Db 236 PheLysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLys 255
QY 2159 TCTCTTCCAAATAAAGCCTTAGAATTTAAAGCAGAGAGAAACACATCAAAAGCAGAGTCTCCT 2218
Db 256 SerValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp 273
QY 2219 GATAATGATGCTCTCTGAAAGCCTACCTGTGGAAGAAAGTTTCTCTTCCAAATAAAGCT 2278
Db 273 ----- 273
QY 2279 TTAGAATTTGAAGGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCAAA 2338
Db 274 ----- 281
QY 2339 CAAAGGATGATGAAGAAATTTCTGGGATTTTCTGGAGTTCCTTTGAGACTCTCTTTCAG 2398
Db 282 GlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeuCysGluThrValSerGln 301
QY 2399 AATGATGTGCTTTTACCCAGGCTACACATCAAAAGAAATTCGATACCTTTAAAGTGAAGA 2458
Db 302 LysAspValCysLeuProLysAla***HisGlnLysGluIleAspLysIleAsnGlyLys 321
QY 2459 TTAGAAGAGTCTCTCGATAAAGATGGTCTTCTGAAAGCCTACCTGTGGAATCAAAATTTCT 2518
Db 322 LeuGluGlySerProValLysAspGlyLeuLysAlaAsnCysGlyMetLysValSer 341
QY 2519 CTTTCCAAATAAAGCCTTAGAATTTGAAGGACAGAGAAACATTCAAAGCAGAGGATGTGAGT 2578
Db 2578 ----- 2578

```

Db	342	IleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGlu	361
QY	2579	TCGTGAGTCACATTCAGTCTTTTGGCAACCCGACTACTGAAATTCACAGTCTACA	2638
Db	362	Lys---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal	376
QY	2639	AAAGTTGAGGAAGACTTAACTCTACTACCAAGGAGGACCAACAAGACAGTAACTGGA	2698
Db	377	-----ProAsnLysAlaLeuGluLeu	383
QY	2699	CAACAGGAAGTGATATTGGCATATTATGAACGAGCTCCACAGATCAAAACAATAAGATG	2758
Db	384	LysAsnGluGlnThrLeuArgAlaAspGluLeuPro-----	396
QY	2759	CCACATCAGCAATAGCAAGAAGAAGATACAAATCACTTCAGATTCGAGATTATC	2818
Db	397	-----SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeu	414
QY	2819	TCTGTGAGTCATACACAGAAATTAGTGTTTACCTGAGGCTACATATCAAAAGAAATA	2878
Db	415	CysGluThrValSerGlnLysAspValCysLeuProLysAla***HisGlnLysGluIle	434
QY	2879	AGACACAAATCGCAAAATAGAGAGTCT-----	2908
Db	435	AspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaPro	454
QY	2908	-----	2908
Db	455	CysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPhe	474
QY	2909	-----CCTGAAAGCCCTTCACATTTGAGCCCTGCCACTGAAATGCAAAACTCT	2956
Db	475	LysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSer	494
QY	2957	GTTCGAATTAAGGCTAGNATGGAGAATAACAAACACTTGAGACA-----	3004
Db	495	ValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPhe	514
QY	3004	-----	3004
Db	515	ProSerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGluSerLeuArg	534
QY	3004	-----	3004
Db	535	GluThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAsp	554
QY	3005	-----GATTCAACTACCTATCAAAATCTTGATGACACTCTCT	3043
Db	555	LysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysLysIleLeuAspThrValHis	574
QY	3044	TCTGTGAAAGAGGAGGAACTTAAAAAGATACTGTGAACAAATTACACCAAAAATG	3103
Db	575	SerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMet	594
QY	3104	GRACAAATGAATAATAGTTTGTGTACTACAAAGGAACCTGCAGAGCGAAGAAATA	3163
Db	595	GluGlnMetLysLysLysPheCysValLeuLysLysLeuSerGluAlaLysGluIle	614
QY	3164	AAATCAGTTGAGAACCAAAAGCTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTG	3223
Db	615	LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerValArgLeu	634
QY	3224	CCTTTAAATCAAGAAGAGAGAGAAATGTCTGATATATATAAAGAAAAAATTAGA	3283
Db	635	ThrLeuAsnGlnGluGluLysArgArgAsnAlaAspIleLeuAsnGluLysIleArg	654
QY	3284	CCC-----GAAGAGCAACTTAGGAAAAAGTTAGAACTGAAACCAACAATT	3328
Db	655	GluGluLeuGlyArgIleGluGluGlnHisArgLysGluLeuGluValLysGlnGlnLeu	674
QY	3329	GAACAGACTCTCAGATACAGATATAGAAATTGAAAGTGTACCAAGTAATTGATCAG	3388
Db	675	GluGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValCysLeuAsnGln	694
QY	3389	GTTTCTCACATCATGAAAGTGAAATGATCTCTTTTCAATGAAATTCATGATTTGAAAAAG	3448
Db	695	ValSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLys	714
QY	3449	GAAATTCATGCTAAACTGGAAGTAGCCACACTGAAACATCAACACCGTGAAGGAA	3508
Db	715	GluIleAlaMetLeuLysLeuGluIleAlaThrLeuLysHisGlnTyrGlnGluLysGlu	734
QY	3509	AATAAATCTTTGAGGACATTAAAGATTTTCAAGAAAGAAATGCTGAACTTCAATGACC	3568
Db	735	AsnLysTyrPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThr	754
QY	3569	CTAAACTGAAACAGAAAACAGTAACAAAAGGCGCTCTCAGTAGAGACGCTTAAA	3628
Db	755	LeuLysLeuLysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGlyGlnLeuLys	774
QY	3629	GTTTTCGCGGAGAGAACACCATGCTGACTTCTTAAATTTGAAGGAA	3673
Db	775	ValLeuIleAlaGluAsnThrMetLeuThrSerLysLeuLysGlu	789
RESULT 6			
US-09-620-405B-494			
; Sequence 494, Application US/09620405B			
; Patent No. 6528054			
; GENERAL INFORMATION:			
; APPLICANT: Jiang, Yugu			
; APPLICANT: Dillon, Davin C.			
; APPLICANT: Mitcham, Jennifer L.			
; APPLICANT: Xu, Jiangchun			
; APPLICANT: Harlocker, Susan L.			
; APPLICANT: Hepler, William T.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND			
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER			
; FILE REFERENCE: 210121.470C8			
; CURRENT APPLICATION NUMBER: US/09/620.405B			
; CURRENT FILING DATE: 2000-07-20			
; NUMBER OF SEQ ID NOS: 495			
; SOFTWARE: FastSEQ for Windows Version 3.0			
; SEQ ID NO 494			
; LENGTH: 743			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: variant			
; LOCATION: (1)...(743)			
; OTHER INFORMATION: Xaa = Any amino acid			
US-09-620-405B-494			
Alignment Scores:			
Pred. No.: 8.49e-135 Length: 743			
Score: 1775.00 Matches: 414			
Percent Similarity: 51.72% Conservative: 68			
Best Local Similarity: 44.42% Mismatches: 137			
Query Match: 27.53% Indels: 313			
DB: 4 Gaps: 13			
US-09-602-362E-26 (1-3673) x US-09-620-405B-494 (1-743)			
QY	950	CTTTTGGAAACATATACGAAATTTACCTAAAAATCTCAAAATACCAATCCAGAAAGAAC	1009
Db	24	LeuLeuGluAsnValIleSerLysThrIleAsnProGlnValSerLysThrGlu-----	41
QY	1010	TCTACAGAACACCTGTAGAGCTGCACCTTGGCGGAAAGAACACCTGCACCGCTGAA	1069
Db	42	-----TyrLys 43	
QY	1070	AGCTTGTCTGGAAGAAACACCTGCAGAGCTGCACGCTTGTGGAGGGAACCTCTGCCAAA	1129
Db	44	GluLeuLeuGlnGluPheIleAspAsnAla-----ThrThrAsnAla 58	
QY	1130	ATTCAATGCTCGGGGAAGCAACATCTCGAAAGTTTGACAGTCAACAGAGAAACACCT	1189

Db 59 IleAspGluLeuLysGluCys-----PheLeuAsnGlnThrAspGluThrLeu 74
QY 1190 AGGAAATT-----TTG 1201
Db 75 SerAsnValGluValPheMetGlnLeuIleTyAspSerSerLeuCysAspLeuPheMet 94
QY 1202 AGGCCTACAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCTAGGAAG 1261
Db 95 SerProAlaLysGluThrSerGluLysPheThrTipAlaAlaLysGlyArgProArgLys 114
QY 1262 ATCCATCGGAGGAGAAAGAAACATCTGTAAGACTGAATCGCTGCAGGAGTTAAACCT 1321
Db 115 IleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThrSer 134
QY 1322 AATAAACTGAAGTTTGGAAAAAGAAACATCTAAATATGATTCATGTCTCTACAAAGAA 1381
Db 135 AsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGlu 154
QY 1382 ACATCTACAAAGCAAGTACAATGTGGATGTGAGTCTGTAGAGCTATATTCAGTCTT 1441
Db 155 SerSerThrLysAlaSerAlaAsn----- 162
QY 1442 TTGGCACACGGACTATTGAAATTCACAGTGTACAAAGTTTCAGGAAGACTTTAATCTT 1501
Db 162 ----- 162
QY 1502 GCTACCAAGATTATCTCTAAGAGTGTGTCACAGAAATTATACGTGTTCCTGATGTACA 1561
Db 162 ----- 162
QY 1562 TATCAAAAGATATCAAAACAATAAATACAAATAGAGATCAGATGTTCCCATCAGAA 1621
Db 163 -----AspGlnArgPheProSerGlu 169
QY 1622 TCCAAACGAGGAGAGATGAAGAAATATCTTGGGATTTCTGGAGTCTCTTTGAGATTCT 1681
Db 170 SerLysGlnGluLysGluLysGluTySerCysAspSerArgSerLeuPheGluSerSer 189
QY 1682 GCAAGACTCAAGTGTGTATACCTGAGTCTATGATATCAGAAAGTAAATGAGATAATAGA 1741
Db 190 AlaLysIleGlnValCysIleProGluSerIleTyGlnLysValMetGluIleAsnArg 209
QY 1742 GAAGTAGAAGAGCTTCTCTGAGAGCCATCTGCCTTCAAGCTCGCGTNGAAATGCAAAAG 1801
Db 210 GluValGluLysProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsn 229
QY 1802 ACTGTTCCAAATAAGCCTTTGAATTGAAGAAATGAACAAACATTCAGAGAGCTCAGATG 1861
Db 230 SerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspProMet 249
QY 1862 TTCCCATCAGATCCAAACAAAGGAGCATGAAGAAATTTCTTGGATTTCTGAGAGTCCC 1921
Db 250 PheProGluSerLysGlnLysAspTyGluGluAsnSerTrpAspSerGluSerLeu 269
QY 1922 TGTGAGACGGTTTCCACAGAGGATGTGTATTATCCAAAGCTACATCAAAAGAAATTC 1981
Db 270 CysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIle 289
QY 1982 GATACCTTAAGTGAATATAGAGAGTCTCCTGTTAAAGATGGTCTCTGAGAGCTTACC 2041
Db 290 AspLysIleAsnGlyLysLeuGluLysProAsnLysAspGlyLeuLeuLysAlaThr 309
QY 2042 TGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAGAGCAGACAGAAACATTC 2101
Db 310 CysGlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThrPhe 329
QY 2102 AAACGAGAGTCTCTGATTAAGATGTGTTCTGAGCCTACCTGTGGAAGGAAAGTTTCT 2161
Db 330 LysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
QY 2162 CTTTCCAAATAAAGCCTTAGAATTAAGAGCAGACAGAAACATCTCAAGACAGAGTCTCCTGAT 2221
Db 350 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp----- 366

QY 2222 AATGATGTCCTTCTTGAAGCCTACCTGTGGAAGGAAAGTTTCTTCTTCCAAATAAAGCTTTA 2281
Db 366 ----- 366
QY 2282 GAATTGAAGGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCTCATCAGATCCAAACAA 2341
Db 367 -----GluIleLeuProSerGluSerLysGln 375
QY 2342 AAGGATGATGAAGAAATTTCTTGGGATTTTGGAGAGTTTCTTGGAGATCTCTTACAGAA 2401
Db 376 LysAspTyGluGluSerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLys 395
QY 2402 GATGTGTTTACCACAGGCTACACATCAAAAGAAATTCGATACCTTAAAGTGGAAATTA 2461
Db 396 AspValCysLeuProLysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeu 415
QY 2462 GAAGAGTCTCTGTATAAGATGGTCTTCTGAAGCCTACTCTGTGGAATGAAATTTCTCTT 2521
Db 416 GluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIle 435
QY 2522 CCAATAAAGCCTTAGAATTGAAGGACAGAGAAACATTCAAAGCAGAGGATGTGAGTTCT 2581
Db 436 ProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLys 455
QY 2582 GTAGAGTCCATTCAGTCTTTTGGCAAAACCGACTACTGAAATTCACAGTCTACAAA 2641
Db 456 ---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal--- 469
QY 2642 GTTGAGGAAGACTTTTAACTTTACTACCAAGGAGGAGCAACAAAGACAGTAACCTGGCAA 2701
Db 470 -----ProAsnLysAlaLeuGluLys 477
QY 2702 CAGNACTGTATATGGCATTTATGAACGAGCTCCACAGATCAACAAATAGATCCCC 2761
Db 478 AsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 489
QY 2762 ACATCAGAAATTAGGAAAAAGAAATACAAATCAACTTCAGATTTCTGAGATTATCTCT 2821
Db 490 ---SerGluSerLysGlnLysAspTyGluGluSerSerTrpAspSerGluSerLeuCys 508
QY 2822 GTGAGTGATACACAGAAATTAAGTGTGTACTGAGGCTACATATCAAAAGAAATTAAG 2881
Db 509 GluThrValSerGlnLysAspValCysLeuProLysAla***HisGlnLysGluIleAsp 528
QY 2882 ACACAAAATGCAAAATAGAGAGTCT----- 2908
Db 529 LysIleAsnGlyLysLeuGluLysProAspAsnAspGlyPheLeuLysAlaProCys 548
QY 2908 ----- 2908
Db 549 ArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLys 568
QY 2909 -----CCTGAAAGCCTTCTCAGCTTGGAGCTGCCTCACTGAAATGCAAAATCTCTGTT 2959
Db 569 AlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 588
QY 2960 CCAATAAAGGCTTAGAATGGAAGAAATAACAAACATTTGAGAGCA----- 3004
Db 599 ProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePro 608
QY 3004 ----- 3004
Db 609 SerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGluSerLeuArgGlu 628
QY 3004 ----- 3004
Db 629 ThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys 648
QY 3005 -----GATTCAACTACCTTATCAAAATCTTGGATCGACTTCTCTCT 3046
Db 649 IleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHisSer 668


```
QY 3047 TGTGAAGAGGAGGAACTTAAAGAAATACTGTGAACAAATTTACAGCAAAATGGAA 3106
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
669 CysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMetGlu 588
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3107 CAAATGAAATAAGTTTGTCTACTACAAAGGAAGCTGTGAGAAAGCAAGAAATAAAA 3166
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
689 GlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluLeuLys 708
QY 3167 TCACAGTTAGAACCAAAAGCTAAATCGGGAACAGAGCTCTGCAGTGTGAGATTGCTT 3226
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
709 SerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerVal 725
QY 3227 TTAATCAAGAAGAGAGAGAGAAATGTCGATATATTAAGAAAAAATTAGACCC 3286
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725 ----- 725
QY 3287 GAAGAGCAACTTAGGAAAAAGTTAGAAGTGAACACCAACTTGAACAGACTCTCAGATA 3346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725 ----- 725
QY 3347 CAAGATATAGAAATTGAAAAGTGTAAACAAGTAATTTGAATCAGGTTTCTCACACTCATGAA 3406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
726 ----- Arg-PheLeuThrLeuMetLy 732
QY 3407 AGTGAATGATCTCTTTCATGAAATTCGATGT 3440
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
732 SmetLysIleSerIleMetLysIleAlaCys 743

RESULT 7
US-09-834-759-494
; Sequence 494, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 494
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(743)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-834-759-494

Alignment Scores:
Pred. No.: 8,49e-135 Length: 743
Score: 1775.00 Matches: 414
Percent Similarity: 51.72% Conservative: 68
Best Local Similarity: 44.42% Mismatches: 137
Query Match: 27.53% Indels: 313
DB: 4 Gaps: 13

US-09-602-362E-26 (1-3673) x US-09-834-759-494 (1-743)

QY 950 CTTTGGACATATACGAAATTAACCTCAAAATACCAATCCAGAGGAACA 1009
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
24 LeuLeuGluAsnValIleSerLysThrIleAsnProGlnValSerLysThrGlu----- 41

QY 1010 TCTACAGGAACACTGATGAGGCTGCACCCCTTGGCGGAGAGACACCTGCACGCGCTGAA 1069
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 42 ----- TyrLys 43
QY 1070 AGCTTGTGTGAAAAAACACCTGACGAGGCTGCAGCTTGTGTGGAGGAACGCTGTGCCAAA 1129
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
44 GluLeuLeuGlnGluPheIleAspAsnAla----- ThrThrAsnAla 58
QY 1130 ATTCATATGCTGGGGAAGCAACATCTGGAAGCTTTGAACAGTCAACAGAGAAACACCT 1189
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 IleAspGluLeuLysGluCys----- PheLeuAsnGlnThrAspGluThrLeu 74
QY 1190 AGGAAAATT----- TTG 1201
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
75 SerAsnValGluValPheMetGlnLeuIleTyrAspSerSerLeuCysAspLeuPheMet 94
QY 1202 AGCCCTACAAAAGAAACATCTGAGAAATTTTCATGCCAGCAAAAGAAAGATCTAGAAG 1261
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 SerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArgLys 114
QY 1262 ATCACATGGGAGGAAAAAGAACATCTGTAAGACATGAATCGTGCAGCAGTAACACCT 1321
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 IleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThrSer 134
QY 1322 AATAAACTGAAGTTTGGAAAAAGAACATCTAATATCATTCATTCCTACAAAGAA 1381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 AsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGlu 154
QY 1382 ACATCTACAAAAGCAAGTACAAATGTGATGTGAGTTCTGTAGACCTATATTCAGTCTT 1441
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 SerSerThrLysAlaSerAlaAsn----- 162
QY 1442 TTTGGCACCGGACTATTGAAAATTCACAGTGTACAAAAGTTGAGGAAGACTTTAATCTT 1501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
162 ----- 162
QY 1502 GCTACCAAGATTATCTCTAAGAGTGTGCACAGAATTATACGTGTTTACCTGATGCTACA 1561
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
162 ----- 162
QY 1562 TATCAAAAAGATATCAAAAATAAATACAAAATAGAAAGATCAGATGTTCCCATCAGAA 1621
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 ----- AspGlnArgPheProSerGlu 169
QY 1622 TCACAAACGAGAGAGAGTGAAGAAATATCTCTGGGATTTCTGGGAGTCTCTTTGAGATCT 1681
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 SerLysGlnGluAspGluLysSerCysAspSerArgSerLeuPheGluSerSer 189
QY 1682 GCAAGACTCAAGTGTGTATACCTCAGTCTATGATCAGAAAGTAAATGAGATAATAGA 1741
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
190 AlaLysIleGlnValCysIleProGluSerIleThrGlnLysValMetGluLeuAsnArg 209
QY 1742 GAAGTAGAAGAGCTTCTCGAAGCCATCTGCCTTCAAGCCTGCCGTGAAATGCAAAAG 1801
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 GluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsn 229
QY 1802 ACTGTTCCAAATAAGCCCTTTGAATTTGAAGTGAACAAACATTTGAGAGCAGCTCAGATG 1861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
230 SerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspProMet 249
QY 1862 TTCCCATCAGAAATCAAAACAAAAGACGATCAAGAAAATTTCTGGGATTTCTGAGAGTCCC 1921
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 PheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSerLeu 269
QY 1922 TGTGAGCGGTTTCACAGAGAGTGTGATTTACCCAAAGCTACACATCAAAAAGAATTC 1981
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 CysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIle 289
QY 1982 GATACCTTAAGTGAAAATTTAGAAAGTCTCCTGTTAAAGATGCTCTCTTGAAGCCCTACC 2041
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 AspLysIleAsnGlyLysLeuGluGluSerProAsnLysaspGlyLeuLeuLysAlaThr 309
QY 2042 TGTGAGAGGAAGTTTCTCTTCCAAATAAGCCCTTAGAATTAAGGACAGAGAAACATTC 2101
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 CysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPhe 329
```

QY 2102 AAAGCAGAGCTCTCTGATTAAGATGCTCTTCAAGCCTACCTGTGGAAGGAAAGTTCT 2161
 Db LysAlaGluProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLysSer 349
 QY 2162 CTTCCAAATTAAGCCTTAGAATTAAGGACAGAAACACTCAAGCAGAGCTCTCTGAT 2221
 Db ValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp 366
 QY 2222 AATGATGGTCTTCTGAAGCCTACCTGTGGAAGGAAAGTTCTCTCCAAATAAGCTTTA 2281
 Db 366
 QY 2282 GAATTGAGCAGAGAAACATTCAAGCAGCTCAGATGTTCCCATCAGATCCAAACAA 2341
 Db 367
 QY 2342 AAGGATGATGAAGAAATTTCTGGATTTTGAGATTTCTTTGAGACTCTCTTACAGAA 2401
 Db LysAspTyrGluGluSerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLys 395
 QY 2402 GATGTGTGTTCCAGGCTACACATCAAAAGAAATTCGATACCTTAAGTGGAAATTA 2461
 Db 396 AspValCysLeuProLysAla**HisGlnLysGluLeuAspLysIleAsnGlyLysLeu 415
 QY 2462 GAAGAGTCTCTGATTAAGATGCTCTTCTGAAGCCTACCTGTGGAATGAAATTTCTCT 2521
 Db 416 GluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyWetLysValSerIle 435
 QY 2522 CCAATTAAGCCTTAGAATTAAGGACAGAAACATTCAAGCAGAGGATGAGTCT 2581
 Db 436 ProThrLysAlaLeuGluMetAspMetGlnThrPheLysAlaGluProGluLys 455
 QY 2582 GTAGTCCACATTCAGTCTTTTGGCAACCGACTACTGAAATTCACAGTCTACAAA 2641
 Db 456 ---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal--- 469
 QY 2642 GTTGAGGAGACTTAACTTACTACCAAGGAGGAGCAACAAACAGATGAGTGGACAA 2701
 Db 470
 QY 2702 CAGGAACGTGATATTGGCATTATTGAACGAGCTCCCAAGATCAACAAATAAGATGCC 2761
 Db 478 AsnGluGlnThrLeuArgAlaAspGluLeuPro----- 489
 QY 2762 ACATCAGATTAAGGAAAGAGATACAAATCAACTCAGATTCAGATTCAGATTCCTCT 2821
 Db 490 ---SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLysCys 509
 QY 2822 GTGAGTGATACAGAAATATAGTGTTCCTGAGGCTACATATCAAAAGAAATAAG 2881
 Db 509 GluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluLeuAsp 528
 QY 2882 ACACAAATGGCAATAGAGTCT----- 2908
 Db 529 LysIleAsnGlyLysLeuGluSerProAspAsnAspGlyPheLeuLysAlaProCys 548
 QY 2908
 Db 549 ArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLys 568
 QY 2909 -----CCTGAAGCCTTCTCATTGAGCCTGCTGCACTGAATGCAAACTCTGTT 2959
 Db 569 AlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 588
 QY 2960 CCAATAAAGCCTTAGAATGAAGAAATAACAAACATTACAGACA----- 3004
 Db 589 ProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePro 608
 QY 3004
 Db 609 SerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGluSerLeuArgGlu 628

QY 3004
 Db 629 ThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys 648
 QY 3005 -----GATTCACTACCTATCAAAAATCTTTGGATGACACTCTCTCT 3046
 Db 649 IleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHisSer 668
 QY 3047 TGTGAAGAGGAGGAACTTAAAGAAAGATACTGTGACAAATTCAGCAAAATCGAA 3106
 Db 669 CysGluArgAlaArgLeuGlnLysAspHisCysGluGlnArgThrGlyLysMetGlu 688
 QY 3107 CAAATCAAAATAAGTTTGTACTACAAAGGAAGTGTGAGAGCGAAGAAATAAAA 3166
 Db 689 GlnMetLysLysPheCysValLeuLysLysSerGluAlaLysGluLys 708
 QY 3167 TCACAGTTAGAGAACCAAAAGCTTAAATGGGAAACAGAGCTCTCTGAGATTCCT 3226
 Db 709 SerGlnLeuGluAsnGlnLysValLysTrpGluGlnLeuLysCysSerVal----- 725
 QY 3227 TTAATCAAGAGAGAGAGAGAAATGTCGATATATTAAAGAAATAATAGACCC 3286
 Db 725
 QY 3287 GAAGACCACTTAGGAAAAAGTTAGAGTGAACACCAACTTGACAGACTCTCAGAATA 3346
 Db 725
 QY 3347 CAAGATATAGATTGAAAGGTAAACAGTAAATTTGAATCAGGTTTCTCACACTCATGAA 3406
 Db 726 -----Arg-PheLeuThrLeuMetLys 732
 QY 3407 ACTGAAATGATCTCTTTTCATGAAAAATTCATGT 3440
 Db 732 sMetLysIleSerTyrMetLysIleAlaCys 743

RESULT 8
 US-09-620-405B-469
 ; Sequence 469, Application US/09620405B
 ; Patent No. 6528054
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepier, William T.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 ; FILE REFERENCE: 210121.470C8
 ; CURRENT APPLICATION NUMBER: US/09/620,405B
 ; CURRENT FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 495
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 469
 ; LENGTH: 650
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(650)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-620-405B-469
 Alignment Scores:
 Pred. No.: 2,24e-134 Length: 650
 Score: 1769.50 Matches: 396
 Percent Similarity: 54.38% Conservative: 57
 Best Local Similarity: 47.54% Mismatches: 112
 Query Match: 27.44% Indels: 268
 DB: 4 Gaps: 9
 US-09-602-362E-26 (1-3673) x US-09-620-405B-469 (1-650)


```
Db 595 GluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluIle 614
QY 3164 AAATCAGATTAGAGAACCAAAAGCTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTG 3223
Db 615 LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluCysSerVal----- 632
QY 3224 CCTTTAAATCAAGAAAGAGAGAGAGAAATGTCGATATATTAAAGAAAAAATTAGA 3283
Db 632 ----- 632
QY 3284 CCGAGAGAGCAACTTAGAAAAAGTTAGAAAGTGAAGTGAACCAACTTGAACAGACTCTCAGA 3343
Db 632 ----- 632
QY 3344 ATACAGATATAGAAATTGAAAAAGTGAACAAGTAATTGAATCAGAGTTTCTCACATCAT 3403
Db 633 -----Arg-PheLeuThrLeuMe 638
QY 3404 GAAAGTGAATGATCTCTTTTCATGAAAAATTGCAATG 3440
Db 638 tLysMetLysIleSerTyrMetLysIleAlaCys 650

RESULT 9
US-09-433-826B-469
; Sequence 469, Application US/09433826B
; Patent No. 6579973
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqi
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.47004
; CURRENT APPLICATION NUMBER: US/09/433,826B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 474
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 469
; LENGTH: 650
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (310)
; NAME/KEY: Xaa = Any Amino Acid<221> unsure
; LOCATION: (429)
; NAME/KEY: Xaa = Any Amino Acid<221> unsure
; LOCATION: (522)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-433-826B-469

Alignment Scores:
Pred. No.: 2,24e-134 Length: 650
Score: 1769,50 Matches: 396
Percent Similarity: 54,38% Conservative: 57
Best Local Similarity: 47,54% Mismatches: 112
Query Match: 27,44% Indels: 268
DB: 4 Gaps: 9

US-09-602-362E-26 (1-3673) x US-09-433-826B-469 (1-650)
QY 1199 TTGAGGCTCAAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCTAGG 1258
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArg 20
QY 1259 AAGATCATCATGGGAGGAAAGAAACATCTGTAAAGACTGAATGCGTGGCAGGAGTAACA 1318
Db 21 LysIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY 1319 CCTAATAAACTGAAGTTTTCGAAAAAGAACATCTAATATGATGTCATGCTCTCAACAAA 1378
```

```
Db 41 SerAsnLysTrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
QY 1379 GAAACATCTACAAAAGCAAGTACAAAATGTGGATGTGAGTTCCTGTAGAGCCTATATTCAGT 1438
Db 61 GluSerSerThrLysAlaSerAlaAsn----- 69
QY 1439 CTTTGTGCACACGAGCTATTGAAAAATTCACAGTGTACAAAAGTTGAGGAGAGACTTTAAT 1498
Db 69 ----- 69
QY 1499 CTTGTCTACCAAGATTATCTCTAAGAGTCTGCACAGAAATTATACGTGTTTACCTGATGCT 1558
Db 69 ----- 69
QY 1559 ACATATCAAAAAGATATCAAAAACAATAATCAAAAATAGAAAGATCAGATGTTCCCATCA 1618
Db 70 -----AspGlnArgPheProSer 75
QY 1619 GAATCCAAAACGAGAGAGAGATGAAGAATATTCTTGGGATTCCTGGAGTCTCTTTGAGAGT 1678
Db 76 GluSerLysGlnGluuAspGluLysSerCysAspSerArgSerLeuPheGluSer 95
QY 1679 TCTGCAAAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAAGTAATCGAGATAAAT 1738
Db 96 SerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsn 115
QY 1739 AGAGAAAGTAGAAGAGCTTCTGAGAAGCCATCTCCCTTCAAGCCTCGCTGNGAAATGCAA 1798
Db 116 ArgGluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln 135
QY 1799 AAGACTGTTCCAAATATAAGCCCTTTGATTTGAAGTAAGTAAGCAACAAATGAGAGAGCTCAG 1858
Db 136 AsnSerValProAsnLysAlaPheGluLysAsnGluGlnThrLeuArgAlaAspPro 155
QY 1859 ATGTTTCCCATCAGAAATCCAAAACAAAGAGCAGTAAAGAAATTTCTGGGATTTCTGAGAGT 1918
Db 156 MetPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTyrAspSerGluSer 175
QY 1919 CCCTGTGAGAGCGTTTCACAGAGAGTGTGATTATCCCAAGCTACACATCAAAAACAA 1978
Db 176 LeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu 195
QY 1979 TTCGATACCTTAAGTGGAAAAATTAGAAGAGTCTCTCTGTTAAAGATGCTTCTGAAAGCT 2038
Db 196 IleAspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAla 215
QY 2039 ACCTGTGGAAGAAAGTTTCTCTCCAAATAAAGCTTAGAATTAAGGACAGAGAAACA 2098
Db 216 ThrCysGlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThr 235
QY 2099 TTCAAAGCAGAGTCTCTCGATAAAGATGGTCTTCTGAAGCCCTACCTGTGGAAGAAAGTT 2158
Db 236 PheLysAlaGluProProGlyLysProSerAlaPheGluProAlaThrGluMetGlnLys 255
QY 2159 TCTCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGACGAGATCTCCT 2218
Db 256 SerValProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAsp----- 273
QY 2219 GATAATGATGCTTCTCTGAAGCCTACCTGTGGAAGAAAGTTTCTCTTCCAAATAAAGCT 2278
Db 273 ----- 273
QY 2279 TTAGAATTGAAGGACAGAGAAACATTCAAGCAGCTCAGATGTTCCCATCAGAAATCCAAA 2338
Db 274 -----GluIleLeuProSerGluSerLys 281
QY 2339 CAAAGGATGATGAAGAAATCTTGGATTTTGGAGTTTCTCTGAGACTCTCTTACAG 2398
Db 282 GlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeuCysGluThrValSerGln 301
QY 2399 AATGATGTGTGTTTACCCCAAGCTACACATCAAAAAGAAATTCGATACCTTAAAGTGGAAAA 2458
Db ----- 2458
```

Db 302 LysAspValCysLeuProLysAla**HisGlnLysGluIleAspLysIleAsnGlyLys 321
QY 2459 TTAGAAGAGCTCTGTATAAGATGCTTCTTGAAGCTACCTGTGGAATGAAAATTCT 2518
Db 322 LeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSer 341
QY 2519 CTTCCAATAAAGCCTTAGAATTAAGAGCAGAGAAACATTCAAAGCAGAGGATGTGAGT 2578
Db 342 IleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProGlu 361
QY 2579 TCTGTAGTCCACATTCAGTCTTTTGGCAACCGACTACTGAAATTCACAGTCTACA 2638
Db 362 Lys---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal 376
QY 2639 AAAGTTGAGGAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAGACAGTAAGTGG 2698
Db 377 -----ProAsnLysAlaLeuGluLeu 383
QY 2699 CAACAGAAAGTGATATGTCATTTATGAACGAGCTCCCAAGATCAACAATAAGATG 2758
Db 384 LysAsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 396
QY 2759 CCCACATCAGAAATTAGCAAGAAAGAGATACAAATCACTTCAGATTCGAGATTATC 2818
Db 397 -----SerGluSerLysGlnLysAspTy:GluGluSerSerTrpAspSerGluSerLeu 414
QY 2819 TCTGTAGTGATACACAGAAATTATGAGTGTTCCTGAGGCTACATATCAAAAGAAATA 2878
Db 415 CysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIle 434
QY 2879 AAGCACAAATGCAAAATAGAGACTCT----- 2908
Db 435 AspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaPro 454
QY 2908 ----- 2908
Db 455 CysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPhe 474
QY 2909 -----CCTGAAAGCCTTCTCACTTTGAGCGTCGCCTGAAATGGAATCTCT 2956
Db 475 LysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSer 494
QY 2957 GTTCCAATAAAGCCTTAGAATGAAGAATAAACAACATTCAGAGCA----- 3004
Db 495 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPhe 514
QY 3004 ----- 3004
Db 515 ProSerGluSerLysGlnLys***ValGluGluAsnSerTrpAspSerGluSerLeuArg 534
QY 3004 ----- 3004
Db 535 GluThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAsp 554
QY 3005 -----GATTCAACTACCTATCAAAAATCTTGGATGCACTCTCT 3043
Db 555 LysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHis 574
QY 3044 TCTTGTGAAGAGAGGAACTTAAAGAAAGTAACCTGTGAAACAATTCAGACAAAATG 3103
Db 575 SerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMet 594
QY 3104 GAACAATGAAAATAAGTTTGTGTACTACAAAGGAAGTGTGAGAGCGAAGAAATA 3163
Db 595 GluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluIle 614
QY 3164 AAATCAGATTAGAACCAACAAAGCTTAATGGGAACAAGAGCTCTGCACTGTGAGATTG 3223
Db 615 LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerVal----- 632
QY 3224 CCTTTAATCAAGAGAGAGAGAGAAATCTCGATATATTAAGAAAGAAATAGA 3283
Db 632 ----- 632

QY 3284 CCCGAGAGCAACTTAGGAAAAAGTTAGAGTGAACACCACTTGAACAGACTCTCAGA 3343
Db 632 ----- 632
QY 3344 ATACAAGATATAGAATTGAAAGTGTAAACAAGTAATTTGAATTCAGTTTCTCACATCAT 3403
Db 633 -----Arg-PheLeuThrLeuMe 638
QY 3404 GAAAGTGAATAATGATCTCTTTTCATGAAATAATTCATGT 3440
Db 638 tlyMetLysIleIleSerTyMetLysIleAlaCys 650
RESULT 10
US-09-604-287A-469
; Sequence 469, Application US/09604287A
; Patent No. 6586572
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepner, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 469
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (310)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: unsure
; LOCATION: (429)
; OTHER INFORMATION: Xaa = Any Amino Acid
; NAME/KEY: unsure
; LOCATION: (522)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-604-287A-469
Alignment Scores:
Pred. No.: 2,24e-134 Length: 650
Score: 1769.50 Matches: 396
Percent Similarity: 54.38% Conservative: 57
Best Local Similarity: 47.54% Mismatches: 112
Query Match: 27.44% Indels: 268
DB: 4 Gaps: 9
US-09-602-362E-26 (1-3673) x US-09-604-287A-469 (1-650)
QY 1199 TTGAGCGCTTACAAAAGAACATCTGAGAAATTTTCATGCCAGCAAAAGAAAGTCTAGG 1258
Db 1 MetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArg 20
QY 1259 AAGATCATCATGGCAGCAAAAACATCTGTAAAGACTGAATCGTGGCAGAGTAACA 1318
Db 21 LysIleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThr 40
QY 1319 CCTATAAACTGAAGTTTGGAAAAAGAACATCTAATATGATGTCATCTCTACAAA 1378
Db 41 SerAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLys 60
QY 1379 GAAACATCTACAAAAGCAAGTACAATCTGATGTGAGTCTCTGTAGACCTATATTACGT 1438
Db 61 GluSerSerThrLysAlaSerAlaAsn----- 69

QY 1439 CTTTTGGCACGGACTATTGAAANTTCAGTGTACAAAAGTTGAGGAGCTTTAAT 1498
Db 69 ----- 69
QY 1499 CTTGCTACCAAGATTATCTCTAGAGTGTCTGCACAGAAATTATACGTGTTTACCTGATGCT 1558
Db 69 ----- 69
QY 1559 ACATATCAAAAGATATCAAAACAATAATCACAAAATAGAGATCAGATGTTCCCATCA 1618
Db 70 -----AspGlnArgPheProSer 75
QY 1619 GAATCCAAACAGAGGAGATGAAGAATATTCTTTGGGATTTCTGGAGTCTCTTTGAGAGT 1678
Db 76 GluSerLysGlnGluGluAspGluLysSerCysAspSerArgSerLeuPheGluSer 95
QY 1679 TCTGCAAGACTCAAGTGTGTATACCTGAGTCTATGTATGATCAGAAAAGTAATGGAGATAAT 1738
Db 96 SerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsn 115
QY 1739 AGAGAGCTAGAGAGCTTCTGAGAGCCATCTGCCTTCAAGCTCGCGTNGAATAATSCAA 1798
Db 116 ArgGluValGluGluProProLysLysProSerAlaPheLysProAlaIleGluMetGln 135
QY 1799 AAGACTGTTCCAAATAAAGCCTTTGAATTGAGAAATGAACAAACATTCAGAGCAGCTCAG 1858
Db 136 AsnSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspPro 155
QY 1859 ATGTTCCCATCAGATCCAAACAAAGACATCAAGAAAATTTCTGGATCTCGAGAT 1918
Db 156 MetPheProGluSerLysGlnLysAspTyrGluGluAsnSerIlePheSerGluSer 175
QY 1919 CCCTGTGACGCTTTCACAGAAGATGTGTATTACCAAAGCTACACATCAAAAAGAA 1978
Db 176 LeuCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGlu 195
QY 1979 TCGATACCTTAAGTGGAAATTAAGAGAGTCTCCTGTGTTAAAGATGTCTTCTGAAGCCT 2038
Db 196 IleAspLysIleAsnGlyLysLeuGluSerProAsnLysAspGlyLeuLeuLysAla 215
QY 2039 ACCTGTGAAGAAAGTTTCTTCCAAATAAAGCCTTAGAATTAAGAGACAGAGAAACA 2098
Db 216 ThrCysGlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThr 235
QY 2099 TTCAAAGCAGAGTCTCTGATATAAGATGCTTCTGAAGCCTACCTGTGGAAGAAAGTT 2158
Db 236 PheLysAlaGluProGluLysProSerAlaPheGluProAlaThrGluMetGlnLys 255
QY 2159 TCTCTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGCAGAGTCTCCT 2218
Db 256 SerValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAsp 273
QY 2219 GATAATGATGCTTCTGAAGCCTACCTGTGGAAGAAAGTTTCTTCTCCAAATAAAGCT 2278
Db 273 ----- 273
QY 2279 TTAGAAATTGAAGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGATCCAAA 2338
Db 274 -----GluIleLeuProSerGluSerLys 281
QY 2339 CAAAGGATGATGAAGAAATTTCTGGGATTTTTCAGAGATTTCTTTCAGACTCTCTTACAG 2398
Db 282 GlnLysAspTyrGluGluSerSerIlePheSerGluSerLeuCysGluThrValSerGln 301
QY 2399 AATGATGTGTTTACCCAGGCTACATCAAAAGAAATTCGATACCTTAAGTGGAA 2458
Db 302 LysAspValCysLeuProLysAla**HisGlnLysGluIleAspLysIleAsnGlyLys 321
QY 2459 TTGAAGAGTCTCCTGATTAAGATGCTTCTGAAGCCTACCTGTGGAATCAAAATTTCT 2518
Db 322 LeuGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSer 341
QY 2519 CTTCCAAATAAAGCCTTAGAATTGAAGACAGAGAAACATTCAAAGCAGAGATGTGAGT 2578

Db 342 IleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGlu 361
QY 2579 TCTGTAGAGTCCACATTTAGTCTTTTGGCAACCGACTACTGAAAATTCAGACTCTACA 2638
Db 362 Lys---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal 376
QY 2639 AAAGTTGAGGAGACTTTAATCTTACTACCAAGGAGGAGCAACAAACACAGACTACTGGA 2698
Db 377 -----ProAsnLysAlaLeuGluLeu 383
QY 2699 CAACAGGAACGTCATATTGGCATATTGAACGAGCTCCACAGATCAAAACAAATAAGATG 2758
Db 384 LysAsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 396
QY 2759 CCCACATCAGATTAGGAAGAAAGACATACAAATCAACTTCAGATTCTGAGATTATC 2818
Db 397 -----SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeu 414
QY 2819 TCTGTGAGTGATACACAGAAATTTATGAGTGTTTTACCTGAGGCTACATATCAAAAAGAAATA 2878
Db 415 CysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluIle 434
QY 2879 AAGACACAAATCGCAAAATAGACAGTCT----- 2908
Db 435 AspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaPro 454
QY 2908 ----- 2908
Db 455 CysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPhe 474
QY 2909 -----CCTGAAAAGCCTTCTCCTTTGAGCCTGCCACTGCAAAATCGAAACTCT 2956
Db 475 LysAlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSer 494
QY 2957 GTTCCAAATAAAGCCTTAGAATGGAAGAAATAAAACAACTTGAGAGCA----- 3004
Db 495 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPhe 514
QY 3004 ----- 3004
Db 515 ProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGluSerLeuArg 534
QY 3004 ----- 3004
Db 535 GluThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAsp 554
QY 3005 -----GATTCAACTACCTTATCAAAAATCTTGGATCGCACTTCT 3043
Db 555 LysIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHis 574
QY 3044 TTTGTGAAGAGGAGGAACTTAAAGAAAGTAAGTAACTGTAACAATACAGCAAAATG 3103
Db 575 SerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMet 594
QY 3104 GAACAAATCAAAAATAAGTTTGTCTACTACAAAAGGAACTCTCAGAAAGCGAAAGAAATA 3163
Db 595 GluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluIle 614
QY 3164 AAATCAGACTTAGAGAACCAAAAAGCTAAATGGGAACCAAGAGCTCTGAGTGTGAGATTG 3223
Db 615 LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuLysSerVal----- 632
QY 3224 CCTTTAAATCAAGAGAGAGAGAGAAATAATGTCGATATATTATAAAGAAAAATTAGA 3283
Db 632 ----- 632
QY 3284 CCGAAGAGCAACTTAGGAAAAAGTTAGAGTGAACCAACCTTGAACAGACTCTCAGA 3343
Db 632 ----- 632
QY 3344 ATACAGATATAGAAATTGAAAAGTGTACCAAGTAATTTGAATCAGGTTTCTCACACTCAT 3403
Db 3403 ----- 3403


```

QY 2639 AAAGTTGAGGAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACCTGGA 2698
DB 377 -----ProAsnLysAlaLeuGluLeu 383
QY 2699 CAACAGGAACGTGATATTGGCATTTATGAACGAGCTCCACAAAGATCAACAAATAAGATG 2758
DB 384 LysAsnGluGlnThrLeuArgAlaAspGluLeuPro----- 396
QY 2759 CCCACATCAGATATTAGGAAGAAAGACATACAAATCAACTTCAGATTCTGAGATTATC 2818
DB 397 -----SerGluSerLysGlnLysAspTyrGluGluSerTrpAspSerGluSerLeu 414
QY 2819 TCTGTGAGTATACACAGAAATATGAGTGTCTTACTGAGGCTACATATCAAAAAGAATA 2878
DB 415 CysGluThrValSerGlnLysAspValCysLeuProLysAla**HisGlnLysGluLe 434
QY 2879 AAGACAACAATGCGAAATAGACAGTCT----- 2908
DB 435 AspLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaPro 454
QY 2908 ----- 2908
DB 455 CysArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPhe 474
QY 2909 -----CCTGAAAGCCTTCTCCTGAGCTGCTGCTGAGCTGCTGAGCTGCTGAGCTGCT 2956
DB 475 LysAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSer 494
QY 2957 GTTCCAAATTAAGGCTTGAAGTGAAGATAAACAACATTCAGAGCA----- 3004
DB 495 ValProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPhe 514
QY 3004 ----- 3004
DB 515 ProSerGluSerLysGlnLys**ValGluGluAsnSerTrpAspSerGluSerLeuArg 534
QY 3004 ----- 3004
DB 535 GluThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAsp 554
QY 3005 -----GATTCAACTACCTTATCAAAAATCTTGGATGCTGCTTCTCT 3043
DB 555 LysIleSerGlyLysLeuGluAspSerThrSerLysSerLysIleLeuAspThrValHis 574
QY 3044 TCTTGTGAAGAGGAGGAACTTAAAGATTAAGTCTGACAAATTTACAGCAAAATG 3103
DB 575 SerCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMet 594
QY 3104 GAACAAATGAAATAAAGTTTGTGTACTACAAAGGAACTGTGAGAGCGGAAAGAAATA 3163
DB 595 GluGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluLe 614
QY 3164 AATACAGATTAGAGAACCAAAAGCTAAATGGGAACAGAGCTCTGAGTGTGAGATTG 3223
DB 615 LysSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerVal----- 632
QY 3224 CCTTTAAATCAAGAGAGAGAGAGAAATGTCGATATATTAAAAAGAAAAAATTAGA 3283
DB 632 ----- 632
QY 3284 CCCGAGAGCAACTTAGGAAAAGTTAGAGTGAAACCAACTTGAACAGACTCTCAGA 3343
DB 632 ----- 632
QY 3344 ATACAAGATATAGATTGAAGAGTCTAACAGTAATTTGAATCAGGTTTCTCACACTCAT 3403
DB 633 -----Arg-PheLeuThrLeuMe 638
QY 3404 GAAAGTGAATATGATCTCTTTCATGAAATATGTCATGT 3440
DB 638 tIysMetLysIleIleSerTyrMetLysIleAlaCys 650

```

RESULT 12

```

US-09-620-405B-473
; Sequence 473, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Repler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 473
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-620-405B-473

```

```

Alignment Scores:
Pred. No.: 8,19e-106 Length: 445
Score: 1414.50 Matches: 315
Percent Similarity: 53.27% Conservative: 43
Best Local Similarity: 46.88% Mismatches: 79
Query Match: 21.94% Indels: 235
DB: 4 Gaps: 6

```

US-09-602-362E-26 (1-3673) x US-09-620-405B-473 (1-445)

```

QY 1568 AAGATATCAAAACAATAAATCAAAAATAGAGATCAGATGTTCCCATCAGATCCAAA 1627
DB 1 LysGluIleAspLysIleAsnGlyLysLeuGluGly----- 12
QY 1628 CGAGAGGAAGATGAAGATATCTTGGGATCTCTGGAGTCTCTTGGAGTCTTCTGAGAGTCTC 1687
DB 13 -----SerProValLysAspGlyLeuLeuLysAlaAsnCysGly 25
QY 1688 ACTCAAGTGTATACCTGAGTCTATGATATCAGAAAGTATGAGATAAATAGAGAAGTA 1747
DB 26 MetLysValSerIleProThrLysAlaLeuGlu-----LeuMetAspMetGlnThrPheLys 44
QY 1748 GAAGACCTTCTCGAGAGCCATCTGCCTTCAAGCCTGCCGTNGAAATGCAAAAGACTGTT 1807
DB 45 AlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 64
QY 1808 CCAATATAAGCCCTTGAATTGAAGATGAACAAACATTTGAGAGCAGCTCAGATGTTCCCA 1867
DB 65 ProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluLeuPro 84
QY 1868 TCAGATCCAAACAAAGGACGATGAGAAATCTTGGGATCTGAGAGTCCCTGTGAG 1927
DB 85 SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeuCysGlu 104
QY 1928 ACGGTTTTCAGAGAGGATGTATTTCACCAAGCTTACATCAACAAAGATTCGATACC 1987
DB 105 ThrValSerGlnLysAspValCysLeuProLysAlaAlaHisGlnLysGluLeuAspLys 124
QY 1988 TTAAGTGGAAATTAAGAGTCTCTGTTAAAGTGTCTTCTGAGCCTACCTGTGGA 2047
DB 125 IleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCysArg 144
QY 2048 AGGAAAGTTCTCTTCCAAATTAAGCCTTTAGAAATTAAGGACAGAGAAACATTCAGCA 2107
DB 145 MetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAla 164
QY 2108 GAGTCTCCTGATAAGATGCTCTTCTGAGAGCTTACTGTGGAGGAGAAAGTTCTCTTCCA 2167
DB 165 GluProGluLysProSerAlaPheGluPro----- 175

```

```
QY 2168 AATAAGCCTTAGAATTAAGACAGACAGAAACACTCAAGCAGAGTCTCCTGATAATGAT 2227
DB 176 -----AlaIleGluMetGlnLys-----181
QY 2228 GGTCTTCTGAAGCCTACCTGTGGAAGGAAGTTTCTTCCAAATAAAGCTTTAGATTG 2287
DB 182 -----SerValProAsnLysAlaLeuGluLeu 190
QY 2288 AAGACACAGAGAAACATTCAAGACGCTCAGATGTTCCCATCAGAAATCCAAACAAAAGGAT 2347
DB 191 LysAsnGluGlnThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 210
QY 2348 GATGAAGAAATCTCTGGGATTTGAGAGTTTCCCTGAGACTCTCTTACAGAAATGATGTG 2407
DB 211 ValGluGluAsnSerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspVal 230
QY 2408 TGTTCACCAAGGTACACATCAAAAGAAATTCGATACCTTAAGTGGAAATATTAGAAGAG 2467
DB 231 CysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu--- 249
QY 2468 TCTCCTGATAAAGATGGTCTTCTGAAGCCTACCTGTGGAAATGAAATTTCTCTTCCAAT 2527
DB 249 -----249
QY 2528 AAGCCTTAGAATTGAAGACAGAGAAACATTCAAGCAGAGGATGTGAGTCTGTAGAG 2587
DB 249 -----249
QY 2588 TCCACATTAGCTCTTTTGGCAACCGACTACTGAAATTTACAGTCTACAAAGTTGAG 2647
DB 249 -----249
QY 2648 GAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAAGTGGACACAGGAA 2707
DB 249 -----249
QY 2708 CGTGATATTGGCATTATTGAACGAGCTCCACAGATCAAAACAAATAAGATGCCACATCA 2767
DB 249 -----249
QY 2768 GAATTAGAGAAAAGAGATACAAAATCAACTTCAGATTCTGAGATTATCTCTGTGAGT 2827
DB 249 -----249
QY 2828 GATACACAGAAATTATGAGTGTGTTTACCTGAGGCTACATATCAAAAGAAATAAAGACAACA 2887
DB 249 -----249
QY 2888 AATGGCAAAATAGAGAGTCTCTGAAAGCCTTCTCACTTTGAGCCTGCCACTGAAATG 2947
DB 249 -----249
QY 2948 CAAACTCTGTTCCAAATAAGGCTTAGAATGGAAGATAAAACAAACATTGAGAGCAGAT 3007
DB 250 -----Asp 250
QY 3008 TCAACTACCCCTATCAAAATCTTGGATGCACTTCTCTTCTGTAAGAGGAGGAACTT 3067
DB 251 SerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeu 270
QY 3068 AAAAAAGATACTGTGACAAATTTACAGCAAAATGGAACMAATGAAAAATAAGTTTGT 3127
DB 271 GlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCys 290
QY 3128 GTACTACAAAAGGAGTGTGAGAGCGAAGAAATAAAATCAACATGATGAGAACCCAAAAA 3187
DB 291 ValLeuLysLysLysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLys 310
QY 3188 GCTAAATGGGAACAGAGCTCTGAGTGTGAGATTGCTTTAAATCAAGAGAGAGAGAG 3247
DB 311 ValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLys 330
```

```
QY 3248 AGAGAAATGTGATATATATTAAGAAAGAAAAATTTAGACCC-----GAAGAG 3292
DB 331 ArgArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGlu 350
QY 3293 CAACCTTAGGAAAAAGTTAGAGTGAACACACCACTTGAACACAGCTCTCAGAAATACAAGAT 3352
DB 351 GlnHisArgGluGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAsp 370
QY 3353 ATAGAATTCAAAAGTGTAAACAGTAATTTGAATCAGGTTTCTCACATCATCAAAAGTGAA 3412
DB 371 IleGluLeuLysSerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGlu 390
QY 3413 AATGATCTCTTTCATCAAAATTCATGTTGTAAGAAAGAAATTCATGCTGCTAAACTCGAA 3472
DB 391 AsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGlu 410
QY 3473 GTAGCCACACTCAAAACATCAACACAGGCTGAAGGAAATAATACTTTGAGACACATTAAG 3532
DB 411 IleAlaThrLeuLysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLys 430
QY 3533 ATTTTCAAGAAAGATGCTGAACCTTCAATGACC 3568
DB 431 IleLeuLysGluLysAsnAlaGluLeuGlnMetThr 442

RESULT 13
US-09-433-826B-473
; Sequence 473, Application US/09433826B
; Patent No. 6579973
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.470C4
; CURRENT APPLICATION NUMBER: US/09/433,826B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 474
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 473
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-433-826B-473

Alignment Scores:
Pred. No.: 8,19e-106 Length: 445
Score: 1414.50 Matches: 315
Percent Similarity: 53.27% Conservative: 43
Best Local Similarity: 46.88% Mismatches: 79
Query Match: 21.94% Indels: 235
DB: Gaps: 6

US-09-602-362E-26 (1-3673) x US-09-433-826B-473 (1-445)
QY 1568 AAGAGATATCAAAACAAATAAATCAAAAATAGAGATCAGATGTTCCCATCAGAAATCCAAA 1627
DB 1 LysGluIleAspLysIleAsnGlyLysLeuGluGly-----12
QY 1628 CGAGAGAGAGATGAAGATATTTCTTGGGATCTCGGAGTCTCTTTGAGAGTTCTCAAG 1687
DB 13 -----SerProValLysAspGlyLeuLeuLysAlaAsnCysGly 25
QY 1688 ACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAAGAAATAGGAGATAATAGAGAAGTA 1747
DB 26 MetLysValSerIleProThrLysAlaLeuGlu---LeuMetAspMetGlnThrPheLys 44
QY 1748 GAAGAGCTTCTCAGAGACCATCTCCCTTCAAGCTGCGGNGAATGCAAAAGACTGTT 1807
DB 45 AlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 64
```

1808 CCAATAAGCCTTTGAATTAAGAAATGAACAAACATTGAGAGCAGCTCAGATGTTCCCA 1867
 65 ProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluLeuPro 84
 1868 TCAGAAATCAACAAAGAGAGATGAAGAAATCTTGGGATTCGTGAGATCCCTGTGAG 1927
 85 SerGluSerLysGlnLysAspTyrGluGluSerTrpAspSerGluSerLeuLysGlu 104
 1928 ACGGTTTCACAGAGGATGTGATTACCCAAAGCTACACATCAAAAGAAATTCGATACC 1987
 105 ThrValSerGlnLysAspValCysLeuProLysAlaAlaHisGlnLysGluLeuAspLys 124
 1988 TTAAGTGGAAATTAAGAGAGCTCTCTGTTAAAGATGGTCTTCTGAAGCCTTACCTGGGA 2047
 125 IleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCysArg 144
 2048 AGGAAGTTTCTCTCCAAATAAAGCCTTAGAATTAAGAGACAGAGAAACATTCAAAGCA 2107
 145 MetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAla 164
 2108 GAGTCTCTGATAAAGATGGTCTCTGAAGCCTACCTGTGGAAGGAAAGTTTCTCTTCCA 2167
 165 GluProProGluLysProSerAlaPheGluPro----- 175
 2168 AATAAGCCTTAGAATTAAGACAGAGAAACACTCAAGCAGAGCTCTCTGATATGAT 2227
 176 -----AlaIleGluMetGlnLys----- 181
 2228 GGTCTCTGAGCCTACCTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCTTTTAGAATTG 2287
 182 -----SerValProAsnLysAlaLeuGluLeu 190
 2288 AAGACAGAGAAACANTCAAGAGCCTCAGATGTTCCCATCAGATCCAAACAAAGAT 2347
 191 LysAsnGluGlnThrLeuAlaAspGlnMetPheProSerGluSerLysGlnLysLys 210
 2348 GATGAAGAAATCTTGGGATTTGAGAGTTTCTCTGAGACTCTTACAGAAATGATGTG 2407
 211 ValGluGluAsnSerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspVal 230
 2408 TGTTTACCAAGGCTACATCAAAAGAAATTCGATACCTTTAAGTGGAAATTAAGAGAG 2467
 231 CysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu--- 249
 2468 TCTCTGATAAGATGGTCTCTGAGCCTACCTGTGGAAGTGAATTTCTCTTCCAAAT 2527
 249 ----- 249
 2528 AAAGCCTTAGAATTAAGACAGAGAAACATTCAAGCAGAGGATGTGAGTTCTGTAGAG 2587
 249 ----- 249
 2588 TCCACATTCAGTCTTTTGGCAACCGACTACTGAAATTCACAGTCTACAAAAGTTGAG 2647
 249 ----- 249
 2648 GAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACTGGACACAGAA 2707
 249 ----- 249
 2708 CGTGATTTGGCATTATTGAACGAGCTCCCAAGATCAAACAATAGATGCCACATCA 2767
 249 ----- 249
 2768 GAATTAGGAGAAAGAGATACAAAATCAACTTCAGATTCGTGAGATTCTCTGTGAGT 2827
 249 ----- 249
 2828 GATACACAGAATTATGAGTGTTTACCTGAGGCTACATATCAAAAGAAATAAGACAACA 2887
 249 ----- 249
 2888 AATGCAAAATAGAGAGTCTCTGAAAGCCTTCTCATTGTGAGCCTGCCACTGAAATG 2947

249 ----- 249
 2948 CAAAACCTCTGTTCCAAATAAAGGCTTTAGAAATGGAAGAATAAAACAAACATTGAGACGAT 3007
 250 -----asp 250
 3008 TCAACTACCTATCAAAATCTTGGATGCACATTCCTTGTGAAAGAGGAAGGAACTTT 3067
 251 SerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeu 270
 3068 AAAAAAGATAACTGTCACAAAATTTACACAAATAAGGAACAATCAAAATAAAGTTTGT 3127
 271 GlnLysAspHisCysGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCys 290
 3128 GTACTACAAAAGAACTGTGAGAGCGAAAGAAATAAATCAAGTTAGAGAACCAAAA 3187
 291 ValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLys 310
 3188 GCTAAATGGGAACAGAGCTCTGAGTGTGAGATTGCCCTTTAAATCAAGAGAGAGAG 3247
 311 ValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLys 330
 3248 AGAAGAAATGTCTGATATATTAAAGAAAAAATTAGACCC-----GAAGAG 3292
 331 ArgArgAsnAlaAspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlu 350
 3293 CACTTAGGAAAAGTTAGAGTGAACACACCACTTGAACAGACTCTCAGATACAGAT 3352
 351 GlnHisArgLysGluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAsp 370
 3353 ATAGATTCGAAAGTGTACAAATTAATTGAATCAGGTTTCTCACACTCATGAAAGTCAA 3412
 371 IleGluLeuLysSerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGlu 390
 3413 AATGATCTCTTTTCATCAAAATTCATGTTTGAAGAAAGAAATTCGATGCTAAACTGAA 3472
 391 AsnTyrLeuHisGluAsnLysMetLeuLysLysGluIleAlaMetLeuLysLeuGlu 410
 3473 GTAGCCACTCAACATCAACACCACTGAGGAGAAATAAATCTTTGAGGACATTAAG 3532
 411 IleAlaThrLeuLysHisGlnTyrGlnGluLysGluAsnLysLysPheGluAspLys 430
 3533 ATTTTACAAAAGAAATGCTCAACTTCAAAATGACC 3568
 431 IleLeuLysGluLysAsnAlaGluLeuGlnMetThr 442

RESULT 14
 US-09-604-287A-473
 ; Sequence 473, Application US/09604287A
 ; Patent No. 6586572
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang Yuqiu
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepler, William T.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
 ; FILE REFERENCE: 21021.470C7
 ; CURRENT APPLICATION NUMBER: US/09/604,287A
 ; CURRENT FILING DATE: 2000-06-22
 ; NUMBER OF SEQ ID NOS: 489
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 473
 ; LENGTH: 445
 ; TYPE: PR1
 ; ORGANISM: Homo sapiens
 ; US-09-604-287A-473
 Alignment Scores: 8.19e-106 Length: 445
 Pred. No.:

Score: 1414.50 Matches: 315
Percent Similarity: 53.27% Conservativity: 43
Best Local Similarity: 46.88% Mismatches: 79
Query Match: 21.94% Indels: 235
DB: 4 Gaps: 6

US-09-602-362E-26 (1-3673) x US-09-604-287A-473 (1-445)

QY 1568 AAAGATATCAAAACATAAATCAAAATAGAGATCAGATGTTCCCATCAGATCCAAA 1627
Db 1 LysGluIleAspLysIleAsnGlyLysLeuGluGly ----- 12

QY 1628 CGAGAGAGAGTGAAGATATCTTGGATCTCGGAGTCTCTTGAGAGTTCTGCAAG 1687
Db 13 -----SerProValLysAspGlyLeuLeuLysAlaAsnCysGly 25

QY 1688 ACTCAAGTGTATACCTGAGTCTATGATCAGAAAGTAATGAGATATAATAGAGAAGTA 1747
Db 26 MetLysValSerIleProThrLysAlaLeuGlu ---LeuMetAspMetGlnThrPheLys 44

QY 1748 GAAGAGCTTCTGAGAGCCATCTGCTTCAAGCTGCGTNGAAATGCAAAAGACTGTT 1807
Db 45 AlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 64

QY 1808 CCAATAAAGCTTTGAAATGAAGATGAACAAACATTGAGAGCAGCTCAGATGTTCCCA 1867
Db 65 ProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeuPro 84

QY 1868 TCAGATCCAAACAAAGGACGATGAAGAAATCTTGGATCTGAGAGTCCCTGTGAG 1927
Db 85 SerGluSerLysGlnLysAspTyrgluLysSerTrpAspSerGluSerLeuLysGlu 104

QY 1928 ACGATTTACAGAGGATGCTATTTACCCAAAGCTACACATCAAAAGAAATTCGATACC 1987
Db 105 ThrValSerGlnLysAspValCysLeuProLysAlaAlaHisGlnLysGluIleAspLys 124

QY 1988 TTAAGTGAATAATAGAGAGTCTCTGTTAAAGATGCTCTCGAAGCCTACCTGTGGA 2047
Db 125 IleAsnGlyLysLeuGluLysProAspAsnAspGlyPheLeuLysAlaProCysArg 144

QY 2048 AGAAAGTTCTTCTCCAAATAAGCCTTAGAATAAGACAGAGAAACATTCAAAGCA 2107
Db 145 MetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAla 164

QY 2108 GAGTCTCTGATAAAGATGGTCTTCTGAAGCTACCTGTGGAAGAAAGTTCTCTTCCA 2167
Db 165 GluProProGluLysProSerAlaPheGluPro ----- 175

QY 2168 AATAAGCCTTAGAATTAAGACAGAGAAACACTCAAGCAGAGTCTCCTGATAATGAT 2227
Db 176 -----AlaIleGluMetGlnLys ----- 181

QY 2228 GGTCTTCTGAAGCCTACCTGTGGAGGAAGTTTCTCTCCAAATAAGCCTTAGAATG 2287
Db 182 -----SerValProAsnLysAlaLeuGluLeu 190

QY 2288 AAGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACAAAGGAT 2347
Db 191 LysAsnGluGlnThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 210

QY 2348 GATGAAGAAATCTTGGATTTTGAGAGTTTCTGAGATCTCTTACAGATGATGTG 2407
Db 211 ValGluGluAsnSerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspVal 230

QY 2408 TGTTTACCAAGGCTACACATCAAAAGAAATTCGATACCTTAAGTGGAAATTAGAAGAG 2467
Db 231 CysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu --- 249

QY 2468 TCTCCTGATAAAGATGGTCTTCTGAAGCCTACCTGTGGAATGAAATTTCTTCCAAAT 2527
Db 249 ----- 249

QY 2528 AAAGCCTTAGAATTTGAAGGACAGAGAAACATTCAAAGCAGAGGATGTGAGTTCTGTAGAG 2587

RESULT 15
US-09-834-759-473

Db 249 ----- 249

QY 2588 TCACACATTGAGTCTTTTGGCAACCGACTACTGAAATTCACAGTCTACAAAGTTGAG 2647

Db 249 ----- 249

QY 2648 GAAGACTTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACTGGCAACAGGAA 2707

Db 249 ----- 249

QY 2708 CGTGATATTGGCATTTATTGAACGAGCTCCACAAGATCAACAAATAAGATGCCACATCA 2767

Db 249 ----- 249

QY 2768 GAATTAGGAAGAAAGAGATACAAATCAACTTCAGATTCTCAGATTATCTCTGTGAGT 2827

Db 249 ----- 249

QY 2828 GATACACAGATTATGAGTGTTCCTGAGGCTACATATCAAAAGAAATAAGACAACA 2887

Db 249 ----- 249

QY 2888 AATGGCAAAATAGAGAGTCTCTGTAAAAGCCTTCTCACTTTTGAGCCTGCCACTGAAATG 2947

Db 249 ----- 249

QY 2948 CAAAACTCTGTTCCAAATAAAGCCTTAGAATGGAAGAATAAAACAACATTGAGAGCAGAT 3007

Db 250 -----Asp 250

QY 3008 TCAACTACCCCTATCAAAATCTTGGATGCACCTCTCTTGTGAAAGAGGAGGAACCTT 3067

Db 251 SerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeu 270

QY 3068 AAAAAGATGAATCTGTGAACAAATATACAGCAAAATATGGAACAAATGAAATATAGTTTGT 3127

Db 271 GlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCys 290

QY 3128 GTACTCAAAAGGAACTGTCAAGAGGAAAGAAATAAATCAAGTGTAGAGAACCAAAA 3187

Db 291 ValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLys 310

QY 3188 GCTAAATGGCAACAGAGCTCTGAGTGTGAGATTGCCTTTAAATCAAGAAAGAGAAG 3247

Db 311 ValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLys 330

QY 3248 AGAAGAAATGTCGATATATATAAAGAAATAATAGACC -----GAAGAG 3292

Db 331 ArgArgAsnAlaAspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlu 350

QY 3293 CAACTTAGCAAAAGTTAGAGTGAACCAACCAACTTGAACACACTCTCAGATACAGAT 3352

Db 351 GlnHisArgLysGluLeuValLysGlnLeuGluGlnAlaLeuArgIleGlnAsp 370

QY 3353 ATAGATTTAAAAGTGTACAAAGTAAATTTGAATCAGGTTTCTCACACTCATGAAAGTAA 3412

Db 371 IleGluLeuLysSerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGlu 390

QY 3413 AATGATCTCTTTCATGAAATTTGCATGTTGAAAGAAATTCCTCATGCTTAAATGAA 3472

Db 391 AsnTyrrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGlu 410

QY 3473 GTAGCCACACTGAAACATCAACCCAGGTGAAGGAAATAAATACTTTGAGACATTAAG 3532

Db 411 IleAlaThrLeuLysHisGlnTyrrGlnGluLysGluAsnLysTyrrPheGluAspIleLys 430

QY 3533 ATTATTCAAGAAAGAAATGCTGAACTTCAATGACC 3568

Db 431 IleLeuLysGluLysAsnAlaGluLeuGlnMetThr 442

; Sequence 473, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 473
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-759-473

Alignment Scores:
Pred. No.: 8,19e-106 Length: 445
Score: 1414.50 Matches: 315
Percent Similarity: 53.27% Conservative: 43
Best Local Similarity: 46.88% Mismatches: 79
Query Match: 21.94% Indels: 235
DB: 4 Gaps: 6

US-09-602-362E-26 (1-3673) x US-09-834-759-473 (1-445)

```
QY 1568 AAAGATATCAAAACAAATTAATCAAAATAGAGATCAGATGTTCCCATCAGAAATCCAAA 1627
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 LysGluIleAspLysIleAsnGlyLysLeuGlu----- 12
QY 1628 CGAGAGGAAGATGAAGATATTTCTGGGATTTCTGGGAGTCTCTTTGAGAGTTCTGCAAG 1687
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 13 -----SerProValLysAspGlyLeuLeuLysAlaAsnCysGly 25
QY 1688 ACTCAAGTGTGATACCTGAGTCTATGATATCAGAAAGTATGAGATATAGAGAAGTA 1747
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 26 MetLysValSerIleProThrLysAlaLeuGlu---LeuMetAspMetGlnThrPheLys 44
QY 1748 GAAGAGCTTCCTGAGAAGCCATCTGCCTCAAGCTGCGGTNGAAATGCAAAAGACTGTT 1807
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 45 AlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVal 64
QY 1808 CCAATAAAGCCTTTGAATTGAAGATCAACAAACATTCAGAGAGCTCAGATGTTCCCA 1867
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGluIleLeuPro 84
QY 1868 TCAGATCCAAACAAAGCAGATCAAGAAATTTCTGGATTTCTGAGAGTCCCTGTGAG 1927
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 SerGluSerLysGlnLysAspTyrGluGluSerIlePaspSerGluSerLeuLysGlu 104
QY 1928 ACGGTTTCACAGAGGATGTGATTTACCCAAAGCTACACATCAAAAGAAATTCGATACC 1987
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 ThrValSerGlnLysAspValCysLeuProLysAlaAlaHisGlnLysGluIleAspLys 124
QY 1988 TTAAGTGGAAATTAAGAGATCTCTGTTAAAGATGCTCTTCTGAGAGCTACCTGTGGA 2047
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 IleAsnGlyLysLeuGluSerProAspAsnAspGlyPheLeuLysAlaProCysArg 144
QY 2048 AGCAAGATTCTCTTCCAAATAAAGCTTAGAATTAAGAGCAGAGAAACATTCAAAGCA 2107
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 MetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAla 164
QY 2108 GAGTCTCTGATGAAGATGCTCTTCTGAGAGCTACCTGTGGAAGAAAGTTCTTCTTCCA 2167
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 GluProProGluLysProSerAlaPheGluPro----- 175
```

```
QY 2168 AATAAAGCCTTAGAATTAAGACGACAGAGAAACACCTCAAGACGAGAGTCTCTCTGATATGAT 2227
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 -----AlaIleGluMetGlnLys----- 181
QY 2228 GGTCTTCTGAAGCCTACCTGTGGAAGAAAGTTTCTCTTCCAAATAAAGCTTTAGAAATG 2287
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 -----SerValProAsnLysAlaLeuGluLeu 190
QY 2288 AAGGACAGAGAAACATTCAAACAGCTCAGATGTTTCCCATCAGATCCAAACAAAGGAT 2347
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 LysAsnGluGlnThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 210
QY 2348 GATGAAGAAATTTCTGGGATTTTCTGAGAGTTCCTTCTGAGACTCTCTTACAGAAATGATG 2407
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 211 ValGluGluAsnSerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspVal 230
QY 2408 TGTTTACCAAGGCTACACATCAAAAGAAATTCGATACCTTAAAGTGGAAATTAAGAGAG 2467
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 CysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu--- 249
QY 2468 TCTCTGTATAAAGATGGTCTTCTGAGAGCTACCTGTGGAATGAAATTTCTCTTCCAAAT 2527
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2528 AAAGCCTTAGAATTCGAAGGACAGAGAAACATTCAAAGCAGAGAGATGTGAGTCTCTAGAG 2587
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2588 TCCACATTCAGTCTTTTGGCAACCGACTACTGAAATTCACACTCTACAAAGTTGAG 2647
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2648 GAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACAGTAACAGCAACAGAA 2707
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2708 CGTGATTTGGCATTTATTGAACGAGCTCCACAGATCAACAAATAGATGCCACATCA 2767
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2768 GAATTAGGAAGAAAGAGAGATACAAATCAAACTTCAGATTTCTGAGATTTATCTCTGAGT 2827
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2828 GATACACAGATTTAGTGTGTTTACCTGAGGCTACATATCAAAAGAAATAAAGCAACA 2887
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2888 AATGGCAAAATAGAGAGTCTCTGAAAGCCTTCTCACTTTGAGCCTGCCACTGAAATG 2947
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 ----- 249
QY 2948 CAAACTCTGTTCCAAATAAAGCTTAGAATGGAAGAAATAAACAACATTGAGACGAGAT 3007
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 -----Asp 250
QY 3008 TCACTACCTTATCAAAATCTTGGTGCACCTTCTCTTGTGAAAGAGGAGGAACTT 3067
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 SerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeu 270
QY 3068 AAAAAAGATAAATCTGCAACAAATTTACAGCAAAATATGGAACAAATGAAATTAAGTTTGT 3127
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 271 GlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCys 290
QY 3128 GTACTCAAAAGAACTGTCAAGCGAAAGAAATAAATCAAGTTAGAGACCAACAAA 3187
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 ValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLys 310
QY 3188 GCTAAATGGGAACAGAGCTCTGAGTGTGAGATGTCCTTTAAATCAAGAAAGAGAGAG 3247
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 311 ValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLys 330
QY 3248 AGAAGAAATGTGATATATATAAAGAAATAATAGACCC-----GAAGAG 3292
```

```
Db      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
331 ArgArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGlu 350
QY      CAACTTAGGAAAAAGTTAGAAAGTGAACACCAACTTGAACAGACTCTCAGAAATACAAGAT 3352
Db      ||||||| :||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 GlnHisArgLysGluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAsp 370
QY      ATAGAATTGAAAAGTGTAAAGTAATTGAATCAGGTTTCTCAGCTCATGAAAGTGAA 3412
Db      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
371 IleGluLeuLysSerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGlu 390
QY      AATGATCTCTTTCATGAAAATTGCATGTTGAAAAGGAAATTGCCATGCTAAACTGAA 3472
Db      ||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
391 AsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGlu 410
QY      GTAGCCCACTGAAACATCAACACCAAGTGAAGGAAAATAATACTTTGAGGACATTAAAG 3532
Db      :||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
411 IleAlaThrLeuLysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLys 430
QY      ATTTTACAGAAAAGAAAGTCACTCAATTCAAATGACC 3568
Db      ||||||| :||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
431 IleLeuLysGluLysAsnAlaGluLeuGlnMetThr 442
```

Search completed: July 15, 2004, 09:29:05
Job time : 94.2381 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:46:47 ; Search time 172.464 Seconds

(without alignments)
13439.296 Million cell updates/sec

Title: US-09-602-362E-26

Perfect score: 6448

Sequence: 1 caagagcttggcgatacaga.....tgactcttaattgaaggaa 3673

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapop 6.0			7.0
Delop 6.0			7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DRV=xlp
-Q/cgn2_1/USFTO_spool_p/US09602362/runat_15072004_093626_22001/app_query.fasta_1.10325
-DB=SPREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=us09602362 @CGN 1.1 814 @runat_15072004_093626_22001 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_25:*

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phage:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5173	80.2	1011	4	Q9bxx2 homo sapien

ID	Q9BXX2	PRELIMINARY:	PRT:	1011 AA.
AC	Q9BXX2;			
DT	01-JUN-2001	(TREMELrel. 17, Created)		
DT	01-JUN-2001	(TREMELrel. 17, Last sequence update)		
DT	01-JUN-2003	(TREMELrel. 24, Last annotation update)		
DE	Breast cancer antigen NY-BR-1.1 (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21174979; PubMed=11280766;			
RA	Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J.,			
RA	Jaeger E., Knuth A., Old L.J., Chen Y.T.;			
RT	"Identification of a Tissue-specific Putative Transcription Factor in Breast Tissue by Serological Screening of a Breast Cancer Library."			
RL	Cancer Res. 61:2055-2061(2001).			
DR	EMBL; AF269088; AAK27326.1; --			
DR	InterPro; IPR002110; ANK.			
DR	Pfam; PF00023; ank. 3.			
DR	SMART; SM00248; ANK. 3.			
DR	PROSITE; PS50088; ANK_REPEAT; 1.			

ALIGNMENTS

RESULT 1

ID	Q9BXX2	PRELIMINARY:	PRT:	1011 AA.
AC	Q9BXX2;			
DT	01-JUN-2001	(TREMELrel. 17, Created)		
DT	01-JUN-2001	(TREMELrel. 17, Last sequence update)		
DT	01-JUN-2003	(TREMELrel. 24, Last annotation update)		
DE	Breast cancer antigen NY-BR-1.1 (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21174979; PubMed=11280766;			
RA	Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J.,			
RA	Jaeger E., Knuth A., Old L.J., Chen Y.T.;			
RT	"Identification of a Tissue-specific Putative Transcription Factor in Breast Tissue by Serological Screening of a Breast Cancer Library."			
RL	Cancer Res. 61:2055-2061(2001).			
DR	EMBL; AF269088; AAK27326.1; --			
DR	InterPro; IPR002110; ANK.			
DR	Pfam; PF00023; ank. 3.			
DR	SMART; SM00248; ANK. 3.			
DR	PROSITE; PS50088; ANK_REPEAT; 1.			


```
QY 2621 GAAATTCACAGTCTACAAAGTTGAGGAGACTTTTAATCTTACTACCAAGGAGGACG 2680
Db 661 GluAsnSerGlnSerThrLysValGluGluAspPheAsnLeuThrThrLysGluGlyAla 680
QY 2681 ACAAGACAGTAAGTGCACACAGGACGATATGGCATTATGACAGAGCTCCACAA 2740
Db 681 ThrLysThrValThrGlyGlnGlnGluArgAspGlyLeuLeuGluArgAlaProGln 700
QY 2741 GATCAAAACAATAAGATGCCCATCATCAGAAATTAGGAAGAAAAGAAGATACAAATCAACT 2800
Db 701 AspGlnThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThr 720
QY 2801 TCAGATCTCAGATTATCTCTGAGTGATGATACACAGAAATATGAGTGTGTACCTGAGGCT 2860
Db 721 SerAspSerGluLeuLeuSerValSerAspThrGlnAsnTyrgluCysLeuProGluAla 740
QY 2861 ACATATCAAAAAGAAAATAAGACAAACAAATGGCAAAATAGACAGTCTCTCTGAAAAGCCT 2920
Db 741 ThrTyrglnLysGluLeuLeuLysThrThrAsnGlnLysIleGlnGluSerProGluLysPro 760
QY 2921 TTCACCTTTGAGCTGCGCATGAAATGCAAAATCTGTTTCCAAATAAAGCTTTAGAAATGG 2980
Db 761 SerHisPheGluProAlaThrGluMetGlnAsnSerValProAsnLysGlyLeuGluTyr 780
QY 2981 AAGAATAAACAACATTTGAGCAGAGATTCACCTACCTATCAAAATCTTGATGCACCTT 3040
Db 781 LysAsnLysGlnThrLeuArgAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeu 800
QY 3041 CCTTCTTTGAAAGAGGAGGAACTTAAAAAAAGATAACTGTGAACAAATTCACAGCAAAA 3100
Db 801 ProSerCysGluArgGlyArgGluLeuLysLysAspAsnCysGluGlnIleThrAlaLys 820
QY 3101 ATGGAACAATGAAATAAGTTTGTGTACTACAAAGAACTGTCAGAGGAAAGAA 3160
Db 821 MetGluGlnMetLysAsnLysPheCysValLeuGlnLysGluLeuSerGluAlaLysGlu 840
QY 3161 ATAAATTCACAGTTAGAGAACCAAAAGCTTAAATGGCAACAGAGCTCTGCAGTGTGAGA 3220
Db 841 IleLysSerGlnLeuGluAsnGlnLysAlaLysTyrGluGlnGluCysSerValArg 860
QY 3221 TTGCTTTTAAATCAAGAAAGAGAGAGAAATGTGATATATTAAGAAAAAATTT 3280
Db 861 LeuProLeuAsnGlnGluGluLysArgArgAsnValAspIleLeuLysGluLysIle 880
QY 3281 AGACCCGACAGCAACTTAGGAAAAGTTAGAGTGAACACCAACTTGAACAGACTCTC 3340
Db 881 ArgProGluGlnLeuArgLysLysLeuGluValLysHisGlnLeuGlnThrLeu 900
QY 3341 AGAATACAGATATAGAAATCAAAAGTGTAAACAGTAATTTGAATCAGGTTTCTCACACT 3400
Db 901 ArgIleGlnAspIleGluLeuLysSerValThrSerAsnLeuAsnGlnValSerHisThr 920
QY 3401 CATGAAGTGAATAATGATCTCTTTCATGAAAATGATGTGAAAAGGAAATGCCATG 3460
Db 921 HisGluSerGluAsnAspLeuPheHisGluAsnCysMetLeuLysLysGluIleAlaMet 940
QY 3461 CTAACCTGGAAGTAGCCACACTGAAACATCAACACACAGGTGAAGGAAATAATACTTT 3520
Db 941 LeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsnLysTyrPhe 960
QY 3521 GAGACATTAAGATTTTACAGAAAAGATGCTGAATCTCAATGACCCCTAAACCTGAAA 3580
Db 961 GluAspIleLysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLys 980
QY 3581 CAGAAAACAGTAACAAAAGGATCTCAGTATAGAGAGAGCTTAAAGTTCTGAGCGCA 3640
Db 981 GlnLysThrValThrLysArgAlaSerGlnTyArgGluGlnLeuLysValLeuThrAla 1000
QY 3641 GAGAACACGATGCTGATCTTAAATTAAGCAA 3673
Db 1001 GluAsnThrMetLeuThrSerLysLeuLysGlu 1011
```

```
RESULT 2
Q9BXX3
ID Q9BXX3 PRELIMINARY; PRT; 1341 AA.
AC Q9BXX3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Breast cancer antigen NY-BR-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=2117479; PubMed=11280766;
RX Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J.,
RA Jaeger E., Knuth A., Old L.J., Chen Y.T.;
RT "Identification of a Tissue-specific Putative Transcription Factor in
RT Breast Tissue by Serological Screening of a Breast Cancer Library.";
RL Cancer Res. 61:2055-2061(2001).
DR EMBL; AF269087; AAK27325.1; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0005515; F:protein binding; NAS.
DR GO; GO:0003700; P:transcription factor activity; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001969; Aspprotease_AS.
DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 1341 AA; 152777 MW; 33E53DDE6FD3A58B CRC64;
Alignment Scores:
Pred. No.: 5,73e-196 Length: 1341
Score: 3458.00 Matches: 740
Percent Similarity: 65.58% Conservative: 83
Best Local Similarity: 58.95% Mismatches: 273
Query Match: 53.63% Indels: 12
DB: Gaps: 4
US-09-602-362E-26 (1-3673) x Q9BXX3 (1-1341)
QY 349 ATGACAGTAGGGAAGAACCCCTCAACCTGAACAAAGAGATATGAAGAGAGGACTGCT 408
Db 1 MetThrLysArgLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArgThrAla 20
QY 409 CTACACTGGCGCTGTGTCAATGGCCCATGCANAAGTAGTAACATTTCTCGTAGACAGAAAG 468
Db 21 LeuHisTyrAlaCysValAsnGlyHisGluGluValValThrPheLeuValAspArgLys 40
QY 469 TCCCGCTTAATGCTTCTGATGGGAGGAGGACACCTCTGATGAAGGCTCTACAAATGC 528
Db 41 CysGlnLeuAspValLeuAspGlyLysHisArgThrProLeuMetLysAlaLeuGlnCys 60
QY 529 GAGAGGGAAGCTTTGTGCAATATTTCTCATAGATGCTGCTGCTATCTAAATTTATGTAGA 588
Db 61 HisGlnGluAla-CysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnLeuValAs 80
QY 589 TGTGTATGGCAACAGCGCTCTCCATTATGCGTTTATAGTGAGATTTATTAAATGCTGCG 648
Db 80 pValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerValValAl 100
QY 649 AACACTGCTGCTCTTGTGTGTCAGTCATCGAGTGCAAAACAAAGGCTAGCCTTCACACCCCT 708
Db 100 aLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuThrProLe 120
QY 709 TTTACTGCCCATACAGAAAAGCAACAACTGTGGAAATTTTACTTAAACAAAATGTC 768
Db 120 uLeuLeuSerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIleLysAsnAl 140
```


Db 736 ---SerGluSerLysGlnLysAspTyrGluGluSerSerTyrAspSerGluSerLeuCy 754
QY 2821 TGTGAGTCATACAGAAATATGAGTGTTCACCTGAGGCTACATATCAAAAAGAAATAAA 2880
Db 754 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluLeuAs 774
QY 2881 GACAAACAATGGCAAAATAGAGAGTCT- 2908
Db 774 pLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCy 794
QY 2908 - 2908
Db 794 sArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLys 814
QY 2909 -CCTGAAAGAGCTTCTCACTTTGAGCTGCCACTGAAATGCAAACTCTGT 2958
Db 814 sAlaGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVa 834
QY 2959 TCCAATAAAGCGCTTAGAATGGAAGATAAACAACATTCAGAGCA- 3004
Db 834 lProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePr 854
QY 3004 - 3004
Db 854 oSerGluSerLysGlnLysValGluGluAsnSerTyrAspSerGluSerLeuArgGlu 874
QY 3004 - 3004
Db 874 uThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys 894
QY 3005 -----GATTCACACTACCTATCAAAATCTTGATGCATCTCTTC 3045
Db 894 sIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHis 914
QY 3046 TTGTGAAGAGGAGGAAGCTTAAAAAGATAACTGTGAACAAATACACCAAAATGGA 3105
Db 914 rCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMetGlu 934
QY 3106 ACAATGAAATAAGTTTGTGTACTACAAAGGAAGTCTCAGAGCGGAAAGAAATAAA 3165
Db 934 uGlnMetLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluLeuLys 954
QY 3166 ATCACACTAGAGAACCAAAAGCTAAATGGGACCAAGAGCTCTGCAGTGTGAGATTGCC 3225
Db 954 sSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluCysSerValArgLeuThr 974
QY 3226 TTTAAATCAAGAGAGAGAGAAATGCGATATATTAATAAGAAATAATAGACC 3285
Db 974 rLeuAsnGlnGluGluLysArgArgAsnAlaAspIleLeuAsnGluLysIleArgGlu 994
QY 3286 C-----GAAGACCACTTAGGAAAAGTTAGAAAGTTAGAGTGAACACCACTTGA 3330
Db 994 uGluLeuGlyArgIleGluGlnHisArgLysGluLeuGluValLysGlnGlnLeuGlu 1014
QY 3331 ACAGACTCTCAGAAATACAGATATAGAAATGAAAGTGAACAAAGTAAATTTGAATCAGGT 3390
Db 1014 uGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGlnVa 1034
QY 3391 TTCTCACATCATGAAGTGAATGATCTCTTTTCATGAAATATGATGTGAAAAGGA 3450
Db 1034 lSerHisThrHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGlu 1054
QY 3451 AATTGCCATGTATAAATCGAAGTAGCCACACTGAAACATCAACACAGGTGAAGGAAA 3510
Db 1054 uIleAlaMetLeuLysLeuGluLeuAlaThrLeuLysHisGlnTyrGlnGluLysGluAs 1074
QY 3511 TAAATCTTTGAGACATTAAGATTTTCAAGAAAAGAAATGCTGAATCAATGACCTC 3570
Db 1074 nLysTyrPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLe 1094
QY 3571 AAAACTCAACAGAAACATACAAAGAGGAGTCTCAGTATAGAGACAGCTTAAAGT 3630
Db 1094 uLysLeuGluGluSerLeuThrLysArgAlaSerGlnTyrSerGlyGlnLeuLysVa 1114

QY 3631 TCTGACGGCAGAGAACACAGATCTGACTTCTAAATTTGAGGAA 3673
Db 1114 lIleuAlaGluAsnThrMetLeuThrSerLysLeuLysGlu 1128
RESULT 3
Q9UPS8
ID PRELIMINARY; PRT; 1709 AA.
AC Q9UPS8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein KIAA1074.
GN KIAA1074.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,
RA Tanaka A., Kozani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205 (1999).
DR EMBL; AB028997; BAA83026.1; -.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00823; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR Hypothetical protein; ANK repeat; Repeat.
SQ SEQUENCE 1709 AA; 196351 MW; ADFA9CDA067120E CRC64;
Alignment Scores:
Pred. No.: 4,366-66 Length: 1709
Score: 1257.00 Matches: 383
Percent Similarity: 43.06% Conservative: 166
Best Local Similarity: 30.04% Mismatches: 323
Query Match: 19.49% Indels: 403
DB: 33 Gaps:
US-09-602-362E-26 (1-3673) x Q9UPS8 (1-1709)
QY 169 GGGGCTGAGCCATCAAGAGGCTCTTAGCTGCGCTGCGCAAGGGCGTGGGGGCCGGAG 228
Db 14 GlySerPheAlaArgArgArgSerSerAlaGlyGly-----GlyGlyGlu 29
QY 229 CCCCCGAACCCCTTCAGGAAGCGGTCTACTAGAAAGGACTACGGGACCATCTACTTC 288
Db 30 Pro-----GlyGluGlyAlaTyrSerGlnProGlyTyr---HisValArgAsp 44
QY 289 GGGGATCTAGGAGAGATCCATACAGTCCCTCCGGGGCCCAAGTCCAGAACTGGAGAG 348
Db 45 ArgAspLeuGlyLysIleHisLysAlaAlaSerAlaGlyAsnValAlaLysValGlnGln 64
QY 349 ATGACAGTAGGAGAGAGCCCGTCAACCTGAACAAAAGAGATATGAAGAGAGGAGCTCT 408
Db 65 lIleLeuLeuLeuArgLysAsn---GlyLeuAsnAspArgAspLysMetAsnArgThrAla 83
QY 409 CTACACTGGGCTGTGTCAATGGCCATGCANAAGTAGTAACTTCTCGTAGACAGAAAG 468
Db 84 LeuHisLeuAlaCysAlaAsnGlyHisProGluValValThrLeuLeuValAspArgLys 103
QY 469 TGCCNGCTTAATGTCTTTGATGGGAGAGGAGGACACCTCTGATGAAGGCTCTCAATGC 528
Db 104 CysGlnLeuAsnValCysAspAsnGluAsnArgThrAlaLeuMetLysAlaValGlnCys 123
QY 529 GAGAGGAGAGCTTGTGCAAAATATCTCATAGATGCTGCTGATCTAAATATGTAGA 588
Db 529 GAGAGGAGAGCTTGTGCAAAATATCTCATAGATGCTGCTGATCTAAATATGTAGA 588

Db 124 GlnGluGluLys-CysAlaThrIleLeuLeuGluHisGlyAlaAspProAsnLeuAlaAs 143
Qy 589 TGTGTATGGCAACACGGCTCTCCATTATGCCGTTTATAGTGAGAAATTTAATATGGTGGC 648
Db 143 pValHisGlyAsnThrAlaLeuHisTyrAlaValTyrAsnGluAspIleSerValAlaTh 163
Qy 649 AACACTGCTGCTTATGGTGCAGTCATCGAGTGCACAAACGCTAGCCCTCACACCCCT 708
Db 163 rLysLeuLeuLeuTyrAspAlaAsnIleGluAlaLysAsnLysAspLeuThrProLe 183
Qy 709 TTTACTGCCATACAGAAAAGCAAGCAAACTGTGGAATTTTACTAAACAAAATGC 768
Db 183 uLeuLeuAlaValSerGlyLysLysGlnMetValGluPheLeuIleLysLysAl 203
Qy 769 AAATGCAACCCATTTAATAGTCTAAATGACACAGCCCTCATGCTGCCATATGTGAGG 828
Db 203 aAsnValAsnAlaValAspLysLeuGluSerSer 214
Qy 829 CTCATCAGAGATAGTCGCGCATGCTCTTTCAGCAAAATGTTGACGCTTTGCTGAACAT 888
Db 214 214
Qy 889 ACATGGAATAACTGCAGAACGTTATGCTGCTGCTGGAGTTAATTACATTCATCAACA 948
Db 215 215
Qy 949 ACTTTTGAACAT--ATACGAAATTTACCTAATAATCCTCAAAATACCAATCCAGAAG 1005
Db 217 uIleSerGluTyrLysGluGluATGTTProLysHis 229
Qy 1006 AACATCTACAGGAACACCTGATGAGCTGCACCCCTGGCGGAAGAACACCTGACAGGC 1065
Db 229 229
Qy 1066 TGAAGCTTGTGGAATAAACACCTGACGAGGCTGCAGGCTTGTGGAGGGAACGTCGC 1125
Db 229 229
Qy 1126 CAAATTCATGCTCTGGGAAAGCAACATCTGAAAGTTTGAACAGTCAACAGAGAAGAAC 1185
Db 229 229
Qy 1186 ACCTAGGAAATTTTGGCCCTACAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAA 1245
Db 229 229
Qy 1246 AGAAGATCTAGGAAGATCACATGGGAGGAAAAAGAACATCTGTAAGACTGAATCGGT 1305
Db 229 229
Qy 1306 GGCAGGAGTAACACCTAATAAACTGAAGTTTTCGAAAAAGGAACATCTAATATGATGC 1365
Db 229 229
Qy 1366 ATGCTCTACAAAGAAACATCTACAAAGCAAGTACAAATCTGGATGTGATTTCTGTAGA 1425
Db 230 SerSerGlnAsnSerAsnSerValAspGluSerSerGluAs 243
Qy 1426 GCTATATTTCAGTCTTTTGGCACAGGACTATTGAAAATTCACAGTGTACAAAGTTGA 1485
Db 243 pSerLeuSerArgLysGlyLysProGlyValAspAspSerTrpProThrSerAspAs 263
Qy 1486 GGAAGACTTTAATCTTGTGTACAGATATCTCTAAG--AG 1524
Db 263 pGluAspLeuAsnPheAspThrLysAsnValProLysProSerLeuAlaLysLeuMetTh 283
Qy 1525 TGCTGCACAGAAATTAATGCTGTTTACCTGATGCTACATATCAAAAGATATCAAAACAT 1584
Db 283 rAlaSerGlnGlnSerArgLysAsnLeuGluAlaThrTyr--GlyThrValArgThrG1 302
Qy 1585 AAATCACAATAA---GAAGATCAGATGTTCCCATCAGATCCAAACAGAGGGAAGATGA 1641
Db 302 yAsnArgThrLeuPheGluAsp-----ArgAspSerAspSe 314

Qy 1642 AGAATATTCTTGGGATTTCTGGAGTCTCTTTGAGAGTTCTCAAGACTCAAGTGTGTAT 1701
Db 314 rGlnAspGluValValGluSerLeuProThrThrSerIleLysValGln---CysPh 333
Qy 1702 ACCTGAGTCTATGTATCAG-----AAAGTAATGGA 1731
Db 333 eSerHisProThrTyrGlnSerProAspLeuProLysProSerHisLysSerLeuAl 353
Qy 1732 GATAAATAGAGAGTAGAGAGTCTCTGAGAAGCCATCTCCCTCAGCCT-----GC 1785
Db 353 aAsnProGlyLeuMetLysGluGluProThrLysProGlyIleAlaLysLysGluAsnG1 373
Qy 1786 CQTNGAAATGCAAAAGACTGTTCCAAATAAAAGCCCTTTGAATTGAAGATAGAACAAACAT 1845
Db 373 yIleAspIleIleGluSerAlaPro-----LeuGluGlnThrAsnAsnAspAsnLe 390
Qy 1846 G-----AGAGCAGCTCAGATGTTCCCATCAGAAATC 1875
Db 390 uThrTyrValAspGluValHisLysAsnAsnArgSerAspMetMetSerAlaLeuGlyLe 410
Qy 1876 CAAACAAAAGGACGATGAAGAAATTTCTGGATTCTGAGAGTCCCTGTGAGACGGTTTC 1935
Db 410 uGlyGlnGluGluAspIleGluSerProTrpAspSerGluSerIleSerGluAsnPhePr 430
Qy 1936 ACAGAAGGATGTATTATCCCAAGACTACATCAAAAGAAATTCGTATACCTTAAGTGG 1995
Db 430 oGlnLysTyrVal-----AspProLeuAlaG1 439
Qy 1996 AAAATTAGAGAGTCTCTGTTTAAAGATGGTCTTCTGAAGCCTACCTGTGAAGGAAAGT 2055
Db 439 yAlaAlaAsp-----GlyLysGluLys 446
Qy 2056 TTCTCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATTCAAAGCAGAGTCTCC 2115
Db 446 sAsnIleGlyAsnGluGlnAlaGlu----- 454
Qy 2116 TGATAAAGATGTCTCTGAGCCCTACCTGTGGAGGAAAGTTTCTCTTCCAAATAAAGC 2175
Db 455 -----AspValPheTyrIleProSerCys---MetSerGlySerArgAsnPheLysMe 471
Qy 2176 CTTAGAATTAAAGGACAGAGAAACACTC-----AAAGCAGAGTCTCTCTGA 2220
Db 471 tAlaLysLeuGluAspThrArgAsnValClyMetProValAlaHisMetGluSerProG1 491
Qy 2221 TAATGATGCTCTCTGAGCCCTACCTGTGGAAAGGAAAGTTTCTCTTCCAAATAAAGCTTT 2280
Db 491 uArgTyrLeuHisLeuLysProThrIleGluMetLysAspSerValProAsnLysAlaG1 511
Qy 2281 AGAATTGAAGGACAGAGAAACATTCAAAGCAGCTCAG-----ATGTTCCC 2325
Db 511 yGlyMetLysAspValGlnThrSerLysAlaAlaGluHisAspLeuGluValAlaSerG1 531
Qy 2326 ATCAGAAATCCAAACAAAAGGATGATGAACAAATTTCTGGGATTTTGAGATTTCTCTGA 2385
Db 531 uGluGluGlnGluArgGlySerGluAsnAsnGlnProGlnValGluGluGluArgLys 551
Qy 2386 GACTCTCTACAGATGATGTGTGTTTACCCAGGCTACACATCAAAAGAAATTCGATAC 2445
Db 551 sLysHisArgAsnAsnGluMetGluValSerAlaAsnIleHis-----AspG1 567
Qy 2446 CTTAAGTGAAATTAGAAGAGTCTCTGATAAAGATGCTCTTCTGAAG----- 2494
Db 567 yAlaThrAspAspAlaGluAspAspAspAspGlyLeuIleGlnLysArgLysSe 587
Qy 2495 -----CCTACCTGCTGAATGAAATTTCTCTCCAATAA 2529
Db 587 rGlyGluThrAspHisGlnGlnPheProArgLysGluAsnLysGluTyrAlaSerGlyPr 607
Qy 2530 AGCCTTAGAATTTGAAGGACAGAGAAACATTCAAAGCAGAGGATGTGAGTCTCTGAGATC 2589
Db 607 oAlaLeuGlnMetLysGluValLysSerThrGluLysGluLysArgThrSerLysGluSe 627

Qy	2590	CACATTCAGT---CTTTTGGCAAAACCGACTACTGTAAGAAATTCACAGTCTACAAAGTTGA	2647
Db	627	rValAsnSerProValPheGlyIysAlaSerLeuLeuThrGlyGlyLeuGlnValAs	647
Qy	2647	GGAAGACTTTAAATCTTACTCAAGAGGAGGACCAACAAAGACACTAACTGGACACAACAGGA	2706
Db	647	pAspAspSerSerLeuSer-----	656
Qy	2707	ACGTGATATTGGCATTAATTGAAACGAGCTCCACAGATCAACAATAAAGATGCCCAATC	2766
Db	656	pGluAspGluGlyArgProThrLysLysThrSerAsnGluLysAsnLysVal-----	673
Qy	2767	AGAATTAGGAAGAAAAGACATACAAATCAACTTCAGAT---TCTCAGATTATCTCTGT	2823
Db	674	-----LysAsnGlnIleGlnSerMetAspValAspAspLeuThrGlnSer	689
Qy	2824	GAGTGATACACAGAAATTATGAGTGT---TTACCTGGAGCTACATATCAAAAAGAAATAAA	2880
Db	689	rSerGluThrAlaSerGluAspCysGluLeuProHisSerSerTyrLys-----	705
Qy	2881	GACAACAAATGGCAAAATAGAGAGTCTCTCGAAAAGCCCTTCACATTTCAGCGCTGCCAC	2940
Db	705	-----	705
Qy	2941	TGAATGCAAAACTCTCTTCCAATAAAGCCTTAGATGGAAGAAATAAACAACATTCGAG	3000
Db	706	-AsnPheMetLeuLeuIleGluGlnLeuGlyMetGluCysLys-----	719
Qy	3001	AGCAGATTCAACTACCCATCAAAAATCTTGGATGCATCTCTTCTGTGAAAGAGGAAG	3060
Db	720	-----AspSerValSerLeuLeuLysIleGlnAspAlaAlaLeuSerCysGluArgLeuLe	738
Qy	3061	GGAACTTAAAAAGATAACTGTCAACAAATACAGCAAAAATGGAACAAATGAAAATAAA	3120
Db	738	uGluLeuLysLysAsnHisCysGluLeuLeuThrValLysIleLysLysMetGluAspLy	758
Qy	3121	GTTTTGTGTACTCAAAAGAACTGTGCAGAACGCAAGAAATAAATCAAGTTAGAGAA	3180
Db	758	sValAsnValLeuGlnArgGluLeuSerGluThrLysGluIleLysSerGlnLeuGluHi	778
Qy	3181	CCAAAAGCTTAAATGGGAACAAGAGCTCTCCAGTGTGAGATTGCCTTTAATCAAGAAGA	3240
Db	778	sGlnLysValGluThrGluArgGluLeuCysSerLeuArgPheSerLeuAsnGlnGluGl	798
Qy	3241	AGAGAAGAGAGAAATGTTCGATATATTAAAGAAAAAAT-----AGACC	3285
Db	798	uGluLysArgAsnAlaAspThrLeuTyrGluLysIleArgGluGlnLeuArgArgLy	818
Qy	3286	CGAAGACCACTTAGGAAAAGTTAGAGTGAACACCACTTCAACAGACTCTCAGAAT	3345
Db	818	sGluGluGlnTyrArgLysGluValGluValLysGlnLeuGluLeuSerLeuGlnTh	838
Qy	3346	ACAAGATATAGAATTGAAAAGTGTCAACAGTAATTTCAATCAGGTT-----	3391
Db	838	rLeuGluMetLeuLeuArgThrValLysSerAsnLeuAsnGlnValValGlnGluArgAs	858
Qy	3391	-----	3391
Db	858	nAspAlaGlnArgGlnLeuSerArgGluGlnAsnAlaArgMetLeuGlnAspGlyIleLe	878
Qy	3391	-----	3391
Db	878	uThrAsnHisLeuSerLysGlnLysGluIleGluMetAlaGlnLysLysMetAsnSerGl	898
Qy	3392	-----TCTCACACTCATGAAAGTGAAATGATCTCTTTCATGAAAATTCATGTTGAAAAA	3447
Db	898	uAsnSerHisSerHisGluGluGluLysAspLeuSerHisLysAsnSerMetLeuGlnGl	918
Qy	3448	GGAATTCGCATGCTAAATCTGAAGTAGCCACACTCAACATCAACACACAGCTGAAGGA	3507
Db	918	uGluIleAlaMetLeuArgLeuGluIleAspThrIleLysAsnGlnAsnGlnGlnLysGl	938
Qy	3508	AAATAAATACTTTTCAGACACATTAAAGATTTTACAAGAAAAAGAAATGCTGAACCTTCAAAATGAC	3567

Db	938	uLyLyCysPheGluAspLeuLysIleValLysGluLysAsnGluAspLeuGlnLysTh	958
Qy	3568	CCTAAATCTGAACAGAAACAGTAACAAAGGCGCATCTCAGTATAGACAGACGCTTAA	3627
Db	958	rLysGlnAsnGluThrLeuThrGlnThrIleSerGlnTyraAsnGlyArgLeuSe	978
Qy	3628	AGTTCTGACGGCAGAGAACAGATGCTGACTCTCTTAATTCGAAG	3670
Db	978	rValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGlu	992
RESULT 4			
Q9H1Q1	PRELIMINARY; PRT; 1710 AA.		
AC	Q9H1Q1		
DT	01-MAR-2001	(TrEMBLrel. 16, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DT	01-CCR-2002	(TrEMBLrel. 22, Last annotation update)	
DE	BAI4528.1	(Hypothetical protein KIAA1074).	
GN	BAI4528.1		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Bird C.;		
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; ALI62272; CACI9649.1; -;		
DR	InterPro; IPR002110; ANK.		
DR	Pfam; PF00023; ank; 4.		
DR	SMART; SM00248; ANK; 4.		
DR	PROSITE; PS50088; ANK_REPEAT; 4.		
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.		
KW	Hypothetical protein; ANK repeat; Repeat.		
SQ	SEQUENCE 1710 AA; 196410 MW; 01CBF9BADB894872 CRC64;		
Alignment Scores:			
Pred. No.:	1,05e-65	Length:	1710
Score:	1250.50	Matches:	382
Percent Similarity:	43.10%	Conservative:	168
Best Local Similarity:	29.94%	Mismatches:	322
Query Match:	19.39%	Indels:	404
DB:	4	Gaps:	34
US-09-602-362B-26 (1-3673) x Q9H1Q1 (1-1710)			
Qy	169	GGGCGTCGACCCATGAAGAGCTT	TAGTCGCGCTGGCAAGGCGTGGCGGCGCCGGAG 228
Db	14	GlySerPheAlaArgGlnArgSer	AlaGlyGly-----GlyGlyGlu 29
Qy	229	CCCCGGAACCCCTTCAGCGAACGGGCT	CACATGAGAAGGACTACGGGACCATCTACTTC 288
Db	30	Pro-----	-----GlyGluGlyAlaTySerGlnProGlyTyr---HisValArgAsp 44
Qy	289	GGGATCTAGGGAAGATCCATACAGCT	GCTCCCGGGGCCAAGTCCAGAGAGCTGGAGAAG 348
Db	45	ArgAspLeuGlyLysIleHisLysAla	AlaSerAlaGlyAsnValAlaLysValGlnGln 64
Qy	349	ATGACATGCGGAAGAACCCCGCTCAAC	CTCGAACAAAGAGATATGAAGAAGAGGACTGCT 408
Db	65	IleLeuLeuLeuArgLysAsn---	GlyLeuAsnAspArgAspLysMetAsnArgThrAla 83
Qy	409	CTACATGGCGCTGTGTCTCAATAGTCC	CGCATGCGCANAGTAGTAACATTTCTGGTAGACAGAAAG 468
Db	84	LeuHisLeuAlaCysAlaAsnGlyHis	ProGluValValThrLeuLeuValAspArgLys 103
Qy	469	TGCNGCTTAATCTCTTGATGCGAAGG	GAGACACCTCTGATGAAGGCTCTACAATGC 528
Db	104	CysGlnLeuAsnValCysAspAsnGlu	AsnArgThrAlaLeuMetLysAlaValGlnCys 123
Qy	529	GAGGGGAAGCTTTGTGCAAAATATTC	TATGATGCTGGTGAATCTAAATATGTAGA 588

Db 124 GlnGluGluLys-CysAlaThrIleLeuLeuGluHisGlyAlaAspProAsnLeuAlaas 143
 QY 589 TGTGTATGGCAACACGGCTCTCCATTATGCGTTATAGTGAGAAATTTAATATGTGGC 648
 Db 143 pValHisGlyAsnThrAlaLeuHisTyrAlaValTyrAsnGluAspIleSerValAlaTh 163
 QY 649 AACACTGCTGCTTCATGGTGGAGTCATCGAGTGCAGAGTGCAGAGTGCAGAGTGC 708
 Db 163 rLysLeuLeuLeuTyrAspAlaAsnIleGluAlaLysAsnLysAspLeuThrProLe 183
 QY 709 TTTACTGCCATACAGAAAGCAAGCAACAACTGTGGAATTTTACTAAACAAAATGC 768
 Db 183 uLeuLeuAlaValSerGlyLysLysGlnMetValGluPheLeuIleLysLysAl 203
 QY 769 AAATGCAACCCATTTAATAGTCTAAATGACACAGCCCTCATGCTGCCATATGTGAGG 828
 Db 203 aAsnValAsnAlaValAspLysLysLeuGluSerSer- 214
 QY 829 CTCATCAGAGATAGTCGGCATGCTCTCTCAGCAAAATGTTGACGCTCTTGTGAGACAT 888
 Db 214 ----- 214
 QY 889 ACATGGAATAACTGCAGAACGTTATGCTGCTGCTGCTGAGTAAATACATTCATCAACA 948
 Db 215 -----HisGlnLe 217
 QY 949 ACTTTTGAACAT--ATACGAAAATACCTTAAATCCTCAAAATACCAATCCAGAGG 1005
 Db 217 uIleSerGluTyrLysGluAlaGlyProLysHis- 229
 QY 1006 AACCTCTACAGGAACACCTGATGAGTGCACCCCTTGGCGGAAAGAACACCTGACAGGC 1065
 Db 229 ----- 229
 QY 1066 TGAAGCTTGCGAAGAAACACCTGACGAGGCTGACGCTTGTGTGAGGAGACGTGTC 1125
 Db 229 ----- 229
 QY 1126 CAAATTCATGTCTGGGAAAGCAACATCTGGAAGTGTGAACAGTCAACAGAGAAAC 1185
 Db 229 ----- 229
 QY 1186 ACCTAGGAAATTTGAGGCTTACAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAA 1245
 Db 229 ----- 229
 QY 1246 AGAAGATCTAGGAAGATCACATGGAGGAAAGAAACATCTGTAAAGACTGAATGCGT 1305
 Db 229 ----- 229
 QY 1306 GGCAGGAGTAACACCTAATAAACTGAAGTTTGGAAAAAGGAACATCTAATATGATGC 1365
 Db 229 ----- 229
 QY 1366 ATGTCCTACAAAGAAACATCTACAAAGCAAGTACAAATGTGGATGTGATCTGTAGA 1425
 Db 230 -----SerSerGlnAsnSerAsnSerValAspGluSerSerGluAs 243
 QY 1426 GCTTATATTCAGTCTTTTGGCACAGGACTATTGAAAATTCACAGTGTACAAAGTTGA 1485
 Db 243 pSerLeuSerArgLeuSerGlyLysProGlyValAspAspSerTrpProThrSerAspAs 263
 QY 1486 GGAAGACTTAACTTGTGTACCAAGATTATCTCTAAG-----AG 1524
 Db 263 pGluAspLeuAsnPheAspThrLysAsnValProLysProSerLeuAlaLysLeuMetTh 283
 QY 1525 TGCTGCACAGAAATATACGCTGTTTACCTGATGCTACATATCAAAAGATATCAAAACAAAT 1584
 Db 283 rAlaSerGlnInSerArgLysAsnLeuGluAlaThrTyr---GlyThrValArgThrGl 302
 QY 1585 AAATCACAATA---GAAGATCAGATGTTCCCATCAGAAATCCAAAGCAGAGGAGTGA 1641
 Db 302 yAsnArgThrLeuPheGluAsp-----ArgAspSerAspSe 314

QY 1642 AGAATATTTCTGGGATTCCTGGAGTCTCTTTGAGAGTTCTGCAAGACTCAAGTGTGTAT 1701
 Db 314 rGlnAspGluValValGluSerLeuProThrThrSerIleLysValGln---CysPh 333
 QY 1702 ACCTGAGCTTATGATCAG-----AAAGTAATGGA 1731
 Db 333 eSerHisProThrTyrGlnSerProAspLeuProLysProSerHisLysSerLeuAl 353
 QY 1732 GATAAATAGAGAGTAGAGAGCTTCCTGAGAACCCATCTCCCTTCAGCCT-----GC 1785
 Db 353 aAsnProGlyLeuMetLysGluProThrLysProGlyIleAlaLysLysGluAsnGl 373
 QY 1786 CQTNGAAATGCAAAAGACTGTTCCAAATAAAAGCCCTTTGAATTGAAGATGAACAAACAT 1845
 Db 373 yIleAspIleIleGluSerAlaPro-----LeuGluGlnThrAsnAsnAspAsnLe 390
 QY 1846 G-----AGAGCAGCTCAGATGTTCCCATCAGATC 1875
 Db 390 uThrTyrValAspGluValHisLysAsnAsnArgSerAspMetSerAlaLeuGlyLe 410
 QY 1876 CAAACAAAGGACGATGAGAAATTTCTGGATCTCAGAGTCCCTCTGAGACGGTTTC 1935
 Db 410 uGlyGlnGluGluAspIleGluSerProTrpAspSerGluSerIleSerGluAsnPhePr 430
 QY 1936 ACAGAAGGATGTGTTATTTACCCAAAGCTACACATCAAAAGAAATTCGATACCTTAAGTGG 1995
 Db 430 oGlnLysTyrVal-----AspProLeuAlaGl 439
 QY 1996 AAAATTAGAGAGTCTCCTGTTTAAAGATGCTTCTGAAGCCTACCTCTGGAAGCAAGT 2055
 Db 439 yAlaAlaAsp-----GlyLysGluLys 446
 QY 2056 TTCTCTTCCAAATAAGCCTTAGAATTAAAGACACAGAGAAACATTCAAAGCAGAGTCTCC 2115
 Db 446 sAsnIleGlyAsnGluGlnAlaGlu----- 454
 QY 2116 TGATAAAGATGCTCTCTGAGGCTACCTGTGGAAGAAAGTTTCTCTTCCCAATAAAGC 2175
 Db 455 -----AspValPheTyrIleProSerCys---MetSerGlySerArgAsnPheLysMe 471
 QY 2176 CTTAGAAATTAAGGACAGAGAAACACTC-----AAAGCAGAGTCTCTCTGA 2220
 Db 471 tAlaLysLeuGluAspThrArgAsnValGlyMetProValAlaHisMetGluSerProGl 491
 QY 2221 TAATGATGCTCTCTGAGGCTACCTGTGGAAGAAAGTTTCTCTTCCAAATAAAGCTTT 2280
 Db 491 uArgTyrLeuHisLeuLysProThrIleGluMetLysAspSerValProAsnLysAlaGl 511
 QY 2281 AGAATTGAGGACAGAGAAACATTCAAAGCAGCTCAG-----ATGTTCCC 2325
 Db 511 yGlyMetLysAspValGlnThrSerLysAlaAlaGluHisAspLeuGluValAlaSerGl 531
 QY 2326 ATCAGAATCCAAACAAAGGATGATGAGAAATTTCTTGGGATTTTGAGAGTTTCTCTGA 2385
 Db 531 uGluGluGlnGluArgGlyLysGluAsnAsnGlnProGlnValGluGluArgLys 551
 QY 2386 GACTCTCTACAGATGATGTGTGTTTACCCAGGCTACACATCAAAAGAAATTCGATAC 2445
 Db 551 sLysHisArgAsnAsnGluMetGluValSerAlaAsnIleHis-----AspGl 567
 QY 2446 CTTAAGTGAAAATTAGAGAGTCTCTGATAAAGATGCTCTTCTGAAGCCTACCTGTGG 2505
 Db 567 yAlaThrAspAspAlaGluAspAspAspAspGlyLeuIleGlnLysArgLysSe 587
 QY 2506 ATGAAAATTT-----TCTCTTCCAAATAAA----- 2530
 Db 587 rGlyGluThrAspHisGlnInPheProArgLysGluAsnLysGluTyrAlaSerSerGl 607
 QY 2531 ----GCCTTAGAATTCAGGACAGAGAAACATTCAAAGCAGAGGATGTGAGTCTGTAGA 2586
 Db 607 yProAlaLeuMetLysGluValLysSerThrGluLysGluLysArgThrSerLysGl 627


```

Db      273 ----- 273
QY      1345 AGGAACATCTAATATGATTGCATGCTCTACAAAAGAAACATCTACAAAAGCAAGTACAAA 1404
Db      273 ----- 273
QY      1405 TGTGGATGTGAGTTCTGTAGAGCCTATATTAGTCTTTTGGCACCGACTATTGAAA 1464
Db      274 ----- 274
QY      1465 TTCACAGTGTACAAAAGTTGAGGAAGACTTTAATCTTCTACCAAGATTTATCTCTAAGAG 1524
Db      284 rLeuProAlaSerAsp---AspLysAspLeuAsnValAlaThrLys----- 298
QY      1525 TGCTGCACAGATATATACGTGTTTACCTGATGCTACATATCAAAAAGATATCAAAACAT 1584
Db      299 ----- 303
QY      1585 AAATCACAATAAGAGATCAGATGTTTCCCATCAGATCCAAACGAGAGGAAGTGAAGA 1644
Db      303 ----- 303
QY      1645 ATATCTTTGGGATCTGGGAGTCTCTTTGAGAGTTCTGCAAGACTCAAGTGTGTATACC 1704
Db      304 ----- 310
QY      1705 TGAGTCTATGTATCAGAAAGTAATGGAGATA--AATAGAACTAGAGAGCTTCTCGA 1761
Db      310 oGlySerSerHisGluLysGlyAsnArgIleValAsnGlyGlnGlyGlyProAla 330
QY      1762 GAAGCCATCTCCCTCAAGCTGCGGNGAAATCAAAAGACTGTTCCAAATAAAGCCTT 1821
Db      330 aLysHisProSerLeuLysProSerThrGluValGluAspProAlaValLysGlyAlaLava 350
QY      1822 TGAATTGAAGATGAACAACATTTGAGAGAGCTCAGATGTTCCATCAGATCCAAACA 1881
Db      350 lGlnArgLysAsnValGlnThrLeuArgAlaGlnAlaLeuProValAlaSerGluG 370
QY      1882 AAAGGACGATCAAGAAAATTTCTGGGATTTCTGAGAGTCCCTGTGAGACGGTTTCACAGAA 1941
Db      370 uGluGlnGluArg----- 374
QY      1942 GGATGTGTAATTTACCAAGCTACATCAAAAGAAATTCGATACCTTAAGTGGAAAT 2001
Db      375 ----- 377
QY      2002 AGAAGAGTCTCCTCTTAAAGATGCTCTTCTGAAGCCTACCTGTGGAAGGAAGTTTCTCT 2061
Db      377 ----- 377
QY      2062 TCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATTCAAAGCAGAGTCTCCTGTATA 2121
Db      377 ----- 377
QY      2122 AGATGGTCTTCTGAAGCCTACCTGTGGAAGGAAGTTTCTCTTCCAAATAAAGCCTTAGA 2181
Db      377 ----- 377
QY      2182 ATTAAGGACAGAGAAACACTCRAAGCAGAGTCTCCTGATAATGATGGTCTTCTGAGACC 2241
Db      377 ----- 377
QY      2242 TACCTGTGGAAGGAAGTTTCTCTTCCAAATAAAGCTTTAGATTGAGGACAGAGAAAC 2301
Db      377 ----- 377
QY      2302 ATTCAAAGCAGCTCAGATGTTCCCATCAGATCCAAACAAAGAGATGATGAAGAAATTC 2361
Db      378 ----- 389
QY      2362 TTGGGATTTTGAGAGTTTCTGAGACTCTCTTACAGAAATGATGTGTGTTTACCCCAAGGC 2421
;

Db      389 n----- 389
QY      2422 TACACATCAAAAAGAAATTCGATACCTTAAGTGAAAATTAGAAGAGTCTCCTGATAAGA 2481
Db      389 ----- 389
QY      2482 TGTCTTCTTGAGCCTACCTACCTGTGGAATGAATAATTTCTTCTTCCAAATAAAGCCTTAGAAT 2541
Db      389 ----- 389
QY      2542 GAAGGACAGAGAAACATTCAAAGCAGGAGTGTGAGTCTGTGTAGATGCCACATTCAGTCT 2601
Db      389 ----- 389
QY      2602 TTTTGGCAACCGACTACTGAAAATTCACAGTCTACAAAGTTGAGGAAGACTTTAATCT 2661
Db      390 ----- 405
QY      2662 TACTACCAAGGGAGCAACAAAGACAGTAAGTAACTGGACACACAGGAACGTGATTTGGCAT 2721
Db      405 pSerThrSerSerAlaIleAlaGlyArgLeuThr-----GlnGlnArgLysIleGly-- 422
QY      2722 TATTGAACGAGCTCCCAAGATCAAAACAAATAAGATGCCCATCAGATCAGAAATTAGGAAGAAA 2781
Db      422 ----- 422
QY      2782 AGHAGATACAAAATCAACTTCAGATTCTCAGATTATCTCTGTGAGTGATACACAGAAATTA 2841
Db      422 ----- 422
QY      2842 TCAGTGTGTTTACCTGAGGCTACATATCAAAAAGAAATTAAGACACAAACAAATGGCAAAATAGA 2901
Db      423 ----- 425
QY      2902 AGAGTCTCTGAAAGCCTTCTCACTTTGAGCGTCCACTGAAATGCAAAATCTGTGTTCC 2961
Db      426 ----- 430
QY      2962 AAATAAGGCTTAGAATGGAAGATAAACAAACATTTGAGCAGATTCAACTACCCCTATC 3021
Db      430 ----- 430
QY      3022 AAAAAATCTTGGATGCACTTCTCTTGTGAAAAGAGAGGAACTTAAAAAGATAACTG 3081
Db      430 ----- 430
QY      3082 TGAACAATTTACAGCAAAAATGGAACAATGAAAATAAGTTTTGTGTACTACAAAAGGA 3141
Db      430 ----- 430
QY      3142 ACTGTCAAGAGCAAGAAATAAAATCACAGTTAGAGAACCAAAAAGCTTAAATGGGAACA 3201
Db      431 ----- 437
QY      3202 AGAGCTCTGCAGTGTGAGATTGCCCTTTTAAATCAAGAAAGAAAGAGAGAGAAATGTGCA 3261
Db      437 sAspArgCysThr-----LeuLysGlnGluAsnGluLysThrAsnValAs 453
QY      3262 TATATTA-----AAAGAAAATTAGACCCGAGAGCAACTTAGGAAAAA 3306
Db      453 nMetLeuTyrLysLysAsnArgGluGluLeuGluArgLysGluLysGlnTyrLysLysG 473
QY      3307 GTTAGAGTGAACCAACTTGAACAGACTCTCTCAGAAATACAGATATAGTAATTGAAAG 3366
Db      473 uValGluAlaLys---GlnLeuGluProThr-----ValGlnSerLeuGluMetLysSe 490
QY      3367 TGTAACA-----AGTAATTTGAATCAGGTTTCTCAGACTCATGAAAGTGAATGATCT 3420
Db      490 rLysThrAlaArgAsnThrProAsnArgAspPheHisAsnHisGluGluMetLysGlyLe 510
QY      3421 CTTTCATGAAAATTCATGTTGAAAAGCAAAATGCCCATGCTATAAATCTGGAAGTAGCCAC 3480
Db      510 uMetAspGluAsnCysIleLeuLysAlaAspIleAlaIleLeuArgGlnGluIleCysTh 530
```



```

Db      381 oLeuGluGlnThrAsnAsnAsp----- 388
Qy      1531 ACAGAATTATACGTGTTTACCTGATCATATCAAAAGAGATATCAAAACATAATATCA 1590
Db      389 -----AsnLeuThrTyValAspGluValHisLysAsnAsnAr 401
Qy      1591 CAAATAGAGATCAGATGTTCCCATCAGAAATCCAAACGAGAGAGATCAAGAAATTC 1650
Db      401 g-----SerAspMetNetSerAlaLeuGlyLeuGlyGlnGluAspGluSerPr 419
Qy      1651 TTGGGATTCCTGGGAGTCTCTTTGAGAGATTCTGCAAGAGACTCAAGTG----- 1696
Db      419 oTTPAspSerGluSerLeuSerGluAsnPheProGlnLysTyValAspProLeuAlaGl 439
Qy      1696 ----- 1696
Db      439 yAlaAlaAspGlyLysGluLysAsnLleGlyAsnGluGlnAlaGluAspValPheTyrl 459
Qy      1697 -----TGATACCTGAGCTATGTATCAGAAAGTA-----ATCGAGATATAATAGAGA 1743
Db      459 eProSerCysMetSerGlySerArgAsnPheLysMetAlaLysLeuGluAspThrArgAs 479
Qy      1744 AGTA-----GAAGAGCTTCTGAGAGCCATCTGCTTCAAGCTGC 1785
Db      479 nValGlyMetProValAlaHisMetGluSerProGluArgTyLeuHisLeuLysProTh 499
Qy      1786 CGTNGAATCAAAAGACTGTTCCAAATAAGCTTTGAATTGAAGAAATCAACAAACATT 1845
Db      499 rIleGluMetCysAspSerValProAsnLysAlaGlyGlyMetCysAspValGlnThrse 519
Qy      1846 GAGAGCAGCT-----CAGATGTTCCCATCAGAAATCCAAACAAAGAGCAGATGA 1893
Db      519 rLysAlaAlaGluHisAspLeuGluValAlaSerGluGluGlnGluArgGluGlyLys 539
Qy      1894 AGAAATTCCT 1903
Db      539 rGluAsnAsn 542

RESULT 9
Q811D2 PRELIMINARY; PRT; 733 AA.
ID Q811D2 PRELIMINARY; PRT; 733 AA.
AC Q811D2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 5730521p14 gene (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II;
RA Strausberg B.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047057; AA047067.1; -.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT NON_TER 733 733
SQ SEQUENCE 733 AA; 80746 MW; 01FE3AAA1F69EA22 CRC64;

Alignment Scores:
Pred. No.: 2,2e-25 Length: 733
Score: 566,50 Matches: 251
Percent Similarity: 36,60% Conservative: 134
Best Local Similarity: 23,86% Mismatches: 275
Query Match: 8,79% Indels: 392
DB: 11 Gaps: 41
US-09-602-362e-26 (1-3673) x Q811D2 (1-733)

```

```

Qy      181 ATGAGAGGCTCTTAGCTGCGCTGGCAAGGGCGTGGCGGGCGCCGAGCCCGGACCC 240
Db      1 MetLysLysIlePheGlyPheArgSerLys-----GlyProSerProLeuGlyPro 17
Qy      241 TTCACGCAACGGGTCTACACTCAGAAAGGACTACCGGACC----- 279
Db      18 SerAlaArgProArgSerAsnCysValGlyPheGlyArgGluSerAlaSerGlySerHis 37
Qy      280 -----ATCTACTTGGGATCTAGGAAGATCCATACAGCTGCCTCCCGG 324
Db      38 ValProArgTyHisIleHisAspLysAspMetGlyLysIleHisLysAlaSerVal 57
Qy      325 GCCCAAGTCCAGAGCTGAGAGATGACAGTA--GGGAAGAGCCCGTCAACCTGAAC 381
Db      58 GlyAspValAlaLysValGlnHisIleLeuLleLeuGlyLysSerGlyVal-----Asn 75
Qy      382 AAAAGAGATATCAAGAGAGAGACTCTCTACACTGGGCTGTGTCAATGGCAATGCANAA 441
Db      76 AspArgAspLysLysAspArgThrAlaLeuHisLeuAlaCysAlaTyArgLysProGlu 95
Qy      442 GTAGTAACATCTTGGTAGACAGAAAGTCCNGCTTAATGTCTTGTGGCGAGGGAGG 501
Db      96 ValValThrLeuLeuValGluArgLysCysGluLleAspAlaArgAspSerGluAsnSer 115
Qy      502 ACACCTCTGATGAAGCTCTACATCGAGAGAGGAGCTTTGTGCAAAATATTCATAGA 561
Db      116 ThrAlaLeuLleLysAlaValGlnCysGlnGluGluGlu-CysAlaAlaLleLeuLys 135
Qy      562 TGCTGCTGCTGATCTAAATTATGTAGATGTGTATGGCAACACGGCTCTCCATTATCGGT 621
Db      135 phiSglyAlaAspProAsnAlaMetAspSerSerGlyAsnThrAlaLeuHisTyAlaVa 155
Qy      622 TTATAGTGAGAAATTTATTAAATGGTGGCAACACTGCTGTCTTCTTGTGTCAGTCATCGAGT 681
Db      155 lTySerGluAsnThrSerMetAlaAlaLysLeuLeuAlaHisAsnAlaAsnLleGluAl 175
Qy      682 GCMAAACAGGCTAGCCTCACACCCCTTTTACTTGCCCATACAGAAAAGAGCAAGCAAC 741
Db      175 aLysAsnLysAspLeuThrProMetLeuLeuAlaValLysAsnLysGlnHisIrl 195
Qy      742 TGTGGAATTTTACTAAACAAAATGCAATGCAACGCATTTAATGATGCTTAATATGCAC 801
Db      195 eValGluPheLeuValLysLysLysAlaSerIleHisAlaValAspGlnLeu----- 212
Qy      802 AGCCCTCATGCTTGCCATATGTGAAGGCTCATCAGAGATAGTCGGCATGCTTCTTCAGCA 861
Db      213 -----GlySer----- 214
Qy      862 AAATGTTGACGCTTTTGTGAAGACATACATGAATAACTGCAGAACGTTATGCTGCTGC 921
Db      214 ----- 214
Qy      922 TCGTGGAGTTAATTACATTCAACAACTTTTGGACACATATACGAAAATTTACCTAAAAA 981
Db      215 -----AsnArgGlnIlePheGluTyArgPheGlyLysArgLeuGlnAr 228
Qy      982 TCCTCAAAATACCAATCCAGAGGAACATCTTACAGGAACACCTGATAGGCTGCACCCCTT 1041
Db      228 gSerGluAsnSerAsnPro-----ValAspAsnGlySerGluAspGlySerLeuThrAr 246
Qy      1042 GCGCGGAAGAACACCTGCACCGCTGAAAGCTTGTGGAAAAAACACCTGACGAGGCTGC 1101
Db      246 gSerTyArgAsnThrProGlyProAlaAspSerTrp-----ProThrSerAspGluGlu-- 263
Qy      1102 ACGCTTGTGGAGGAGACGCTGTCGCAAAATTCATGTCTGGGGAAGCAACATCTGGAAA 1161
Db      264 -----As 264
Qy      1162 GTTTGAACAGTCAACAGAGAACAACCTAGGAAAATTTTGGCCCTCACAAAGAAACATC 1221
Db      264 pTyArgAsnPheAspAsnLysAsnValProLysIleAsnLeuThrGluLeu----- 280

```


DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ25822.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Mueashino K., Yuuki H., Hara H., Suzuki Y., Fukuda S.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RA "NRDO human cDNA sequencing project";
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
EL EMBL; AK096888; BAC05379.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 168 AA; 18305 MW; 8C6BA50DB58944CF CRC64;

Alignment Scores:
Pred. No.: 1,17e-24 Length: 168
Score: 553.00 Matches: 113
Percent Similarity: 79.74% Conservative: 9
Best Local Similarity: 73.86% Mismatches: 27
Query Match: 8.58% Indels: 4
DB: 4 Gaps: 1

US-09-602-362E-26 (1-3673) x Q9N7B8 (1-168)
Qy 1064 GCTGAAGCTTCTGGGAAAAACCTGACAGGCTGACGCTGGTGGAGGGAACGCTCT 1123
Db 2 AlagluSerLeuValGluLysProProAspGluAlaAlaProLeuValGluGlyThrAla 21
Qy 1124 GCCAAATTCATCTCTGGGAAAAGCAACATCTGGAAGTTGGAACAGTCAACACAGAGAA 1183
Db 22 AspLysLeuGlnCysLeuGlyAlaThrSerGlyLysPheGluInSerAlaGluGlu 41
Qy 1184 ACACCTAGGAAATTTTGGAGCCCTACAAAAGAAACATCTGAGAAATTTTCATGCCAGCA 1243
Db 42 ThrProLysLysIleMetArgThrAlaLysGluThrSerGluLysPheAlaTrpProAla 61
Qy 1244 AAAGAAAGATCTAGAGATCAGATCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1303
Db 62 LysGluArgProArgLysIleThrTrpGluLysGluThrSerValLysThrGluCys 81
Qy 1304 GTGCGAGGAGTAACACCTATAAATCTGAAAGTTTGGAAAAAGAAACATCTAATATGATT 1363
Db 82 ValAlaGlyValIleProAsnLysThrGluValLeuGluLysGlyThrSerGluMetLeu 101
Qy 1364 GCATGCTCTACAAAAGAAACATCTACAAAAGCAAGTACAAATGTGGATGTGAGTCTGTGA 1423
Db 102 ThrCysProThrLysGluThrSerThrLysAlaSerThrAsnValAspValSerVal 121
Qy 1424 GAGCTATATTCTAGTCTTTT-----GGCACGGACTATTGAAATTCACAG 1471
Db 122 GluSerIlePheArgValSerProCysHisProGlyTrpAsnAlaMetAlaArgSerArg 141
Qy 1472 TGTACAAAGTTGAGGAAGACTTTAATCTTGTCTACCAAG 1510
Db 142 LeuThrAlaThrSerThrSerGlnLysGluAlaThrLys 154

RESULT 11
Q9CS61
ID Q9CS61 PRELIMINARY; PRT; 533 AA.
AC Q9CS61;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 5730521P14Rik protein (Fragment).
GN 5730521P14Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Matsuda Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Okazaki H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection";
RL EMBL; AK017783; BAB30930.1; -.
DR HSSP; P80144; 2MYO.
DR MCD; MG1:1917887; 5730521P14Rik.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Repeat. 533
FT NON_TER 533
SQ SEQUENCE 533 AA; 58545 MW; FDB7E9116778430 CRC64;

Alignment Scores:
Pred. No.: 5,24e-22 Length: 533
Score: 509.00 Matches: 191
Percent Similarity: 37.03% Conservative: 82
Best Local Similarity: 25.95% Mismatches: 180
Query Match: 7.89% Indels: 283
DB: 11 Gaps: 26

US-09-602-362E-26 (1-3673) x Q9CS61 (1-533)
Qy 181 ATGAAGAGCTCTTAGCTGCCCTGGCAGGCGCTGCGGGCCCGGAGCCCGAACCC 240
Db 1 MetLysLysIlePheGlyPheArgSerLys-----GlyProSerProLeuGlyPro 17
Qy 241 TTTACGCAACGGGTCTACACTGAGAAGGACTACGGGACC----- 279
Db 18 SerAlaArgProArgSerAsnCysValGlyPheGlyArgGluSerAlaSerGlySerHis 37
Qy 280 -----ATCTACTTCGGGATCTAGGAGATCCATACAGCTCCCTCCCG 324
Db 38 ValProArgTyHisIleHisAspLysAspMetGlyLysIleHisLysAlaAlaSerVal 57
Qy 325 GGCCCAAGTCAGAGCTGGAGAGATGACAGTA---GGGAAGACCCCGTCAACCTGAAC 381
Db 58 GlyAspValAlaLysValGlnHisIleLeuIleLeuGlyLysSerGlyVal-----Asn 75
Qy 382 AAAAGAGATATGAAGAAGGAGCTCTCTACATCGGGCTGTGTCAATGGCCATCANAA 441
Db 76 AspArgAspLysLysAspArgThrAlaLeuHisLeuAlaCysAlaTyGlyHisProGlu 95
Qy 442 GTAGTAACATTTCTGAGACAGAGGTCGCGCTTAATGTCTTGTGAGGCGAAGGGAGG 501
Db 96 ValValThrLeuLeuValGluArgLysCysGluIleAspAlaArgAspSerGluSer 115
Qy 502 ACACCTCTGATGAGGCTCTACATGCGAGGAGGAGCTTTGTGCAAAATTTCTCATAGA 561

```

Db      116 ThrAlaLeuIleLysAlaValGlnCysGlnIuGlu-CysAlaAlaIleLeuLeuAs 135
QY      562 TGCTGGTCTGATCTAAATTATAGATGTATGGCAACACGGCTCTCCATTATGCGGT 621
Db      135 pHisGlyAlaAspProAsnValMetAspSerSerGlyAsnThrAlaLeuHisTyrAlaVa 155
QY      622 TTATAGTCAGATTTATTAAATGGTGGCAACTGCTGCTCTATGCTCAGTCATCGAGGT 681
Db      155 ltyrSerGluAsnThrSerMetAlaAlaLysLeuLeuAlaHisAsnAlaAsnIleGluAl 175
QY      682 GCAAAACAGGCTAGCTCACACCCCTTTTACTGCCCATACAGAAAGAAAGCAAGCAAAC 741
Db      175 aLysAsnLysAspAspLeuThrProMetLeuLeuAlaValLysGluAsnLysGlnHisI 195
QY      742 TGTGGAATTTTACTAACAATAATGCAATGCAATGCAATTTAATGAGTCTAAATGCAC 801
Db      195 eValGluPheLeuValLysLysLysAlaSerIleHisAlaValAspGlnLeu 212
QY      802 AGCCCTCATGCTTGCCATATGTAAGGCTCATCAGAGATAGTCGGCATGCTTCTTCAGCA 861
Db      213 -----GlySer----- 214
QY      862 AAATGTTGACGCTTTGCTGAAGACATACATGGAATTAATCCAGAACGTTATGCTGTGTC 921
Db      214 ----- 214
QY      922 TCGTGGAGTTAATTACATTTCATCAACAACTTTTGGACATATACGAAATTTACCTAAAAA 981
Db      215 -----AsnArgGlnMetPheGluTyrAspGlyLysArgLeuGlnAr 228
QY      982 TCTCAAAATACCAATCCAGAGGAACATCTACAGGAACACCTGATGAGGCTGACCCCTT 1041
Db      228 gSerGluAsnSerAsnPro-----ValAspAsnGlySerGluAspGlySerLeuThrAr 246
QY      1042 GCGCAAGAGACACCTGACAGCGCTGGAAGCTTCTGGAAGAAACACCTGACGAGGCTGC 1101
Db      246 sSerTyrAsnThrProGlyProAlaAspSerTrp-----ProThrSerAspGluGlu-- 263
QY      1102 AGCCTTGCTGGAGGAAGCTCTGCCAAAATTCATATGCTCGGGAAGAACACATCTGAAA 1161
Db      264 -----As 264
QY      1162 GTTTGAACAGTCAACAGAGAAACACCTAGGAATAATTTTGAGGCTACAAAGAAACATC 1221
Db      264 pTyrAsnPheAspAsnLysAsnValProLysIleAsnLeuThrGluLeu----- 280
QY      1222 TGAGAAATTTTCATGCGCCAGAAAAGAAAGATCTAGGAAGATCATCTGGGAGGAAAAAGA 1281
Db      281 -----TrpThrAlaAlaGlnGlnSerArgLys----- 289
QY      1282 AACATCTGTAAGACTGAATGCTGGCGAGGAGTACACCTAATAAACTGAAGTTTGGGA 1341
Db      290 -----AsnGlnThrLysCys-----GlyPheGluGluLeuAs 300
QY      1342 AAAAGGAACATCTAATATGATGTCATGCTGCTACAAAGAAACATCTACAAAGCAAGTAC 1401
Db      300 pAsnGlyAla-----ArgPheAspAspSerAspSerProSerG 313
QY      1402 AAATGTTGAGTGTGAGTCTGTAGACCTTATATTCAGTCTTTTGGCACACGCACTATGA 1461
Db      313 uSerGluAspAlaIleGluValGluProAlaProSerVal---ArgValGlnThrLeuSe 332
QY      1462 AAATTCACAGTGTACAAAGTTGACGAGACTTTAATCTGCTACCAAGATTATCTCTAA 1521
Db      332 rProSerArgGlnSer----- 337
QY      1522 GAGTGTGTCACAGAATTATACGTGTTTACCTGATGCTACATATCAAAAGAAATATCAAAAC 1581
Db      338 -----ProAsp----- 339
QY      1582 AATAAATCAAAAATAGAGATCAGATGTTCCCTCAGAAATCCAAACAGAGAG----- 1633

```

```

Db      340 -----ProValGluGlyAlaThrGluLeuAlaI 349
QY      1634 -CAAGATGAAGAATATTTCTTGGGATTTCTGGGAGTCTCTTTGAGAGTCTTGCAGAACTCA 1692
Db      349 eGluGlyGluGluAsnGlyThrAsp-----ValIleGluSerAlaSerGlnGluG 366
QY      1693 AGTGTGTATACCTGAGTCTATGTATCAGAAAGTAATGAGATAAATAGAGAAGTAGAAGA 1752
Db      366 ----- 366
QY      1753 GCTTCTCAGAGCCATCTGCCTTCAAGCCTGCCGTGCAAAATGCAAAAGACTGTTCCAAA 1812
Db      366 ----- 366
QY      1813 TAAAGCCTTTGAATTGAAGAATGAACAAACATTGAGACAGCTCAGATGTTCCCATCAGA 1872
Db      367 -----ProAsnHisAspAsnLeuThrArgAlaAspGlyTyrHisLysSe 381
QY      1873 ATCCAAA-----CAAAAGGACGATGAAGAAATTC 1902
Db      381 rAsnLysSerGluMetMetSerAlaLeuGlyLeuGlyAspGluAspGluHisSerPr 401
QY      1903 TTGGGATTTCTGAGAGTCCCTGTGAGACGCTTTTCACAGAGGATGTG----- 1948
Db      401 cTrpAspSerGluSerIleSerGluSerValSerLeuLysAspValGlyHisPheSerGl 421
QY      1949 -----TATTT 1953
Db      421 yThrAlaAspGlnThrGlyLysArgAlaHisGlyGlnIleGluAspValThrTyrI 441
QY      1954 ACCCAAGCTACACATCAAAAGAAATTCATACCTTAAAGTGA----- 1996
Db      441 eProSerCys-----MetSerGlySerArgSerPheLysMe 453
QY      1997 ----AAATTAGAGAGTCTCTGTTTAAAGATGCTTCTTGAAGCCTACCTGTGGAGGAA 2052
Db      453 tAlaLysLeuGluGluSer-----ArgAs 461
QY      2053 AGTTTCTCTTCCAAATAAGCCCTTAGATTAAAGACAGAGAAACATTCAAGACAGAGTC 2112
Db      461 nValGlyLeuProValAlaHisMetGlyMetGlu-----LysAsnGluAl 476
QY      2113 TCTGTATAAGATGCTTCTTGAAGCCTACCTGTGGAAGGAAAGTTCTCTTCCAAATAA 2172
Db      476 aProArgLysTyrAlaIleMetGluProThrIleGluArgAlaProValLeuAsnly 496
QY      2173 A-----GCCTTAGATTAAAGACAGAGAAACACTCAAGACAGAG 2212
Db      496 sThrGluThrValGlyMetThrAspAlaGlnThrPheLysSerGlu 511

RESULT 12
QH560 PRELIMINARY; PRT; 264 AA.
AC QH560;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE BA526D8.2 (Novel protein similar to KIAA1074).
GN BA526D8.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A. i
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALI36981; CAC12726.1; -.
DR HSSP; Q00420; 1AWC.
DR InterPro; IPR02110; ANK.
DR Pfam; PF00023; ank; 5.
DR SMART; SMC0248; ank; 5.
DR PROSITE; PS50088; ANK_REPEAT; 4.

```

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 264 AA; 30436 MW; AB9237C94CAC0405 CRC64;

Alignment Scores:
Pred. No.: 5.19e-22 Length: 264
Score: 508.50 Matches: 111
Percent Similarity: 62.50% Conservative: 44
Best Local Similarity: 44.76% Mismatches: 85
Query Match: 7.89% Indels: 8
DB: 4 Gaps: 2

US-09-602-362E-26 (1-3673) x Q9H560 (1-264)

QY 241 TTCAGCAACGGGCTCTACATGAGAGACTACGGGACCATCTACCTGGGGGATCTAGG 300
DB 23 TyrAlaGlyArgGlyTyrHisIleArgSerTrp-----GluLeuArg 36
QY 301 AAGATCCATACAGCTGCTCCGGGGCCAAAGTCCAGAGCTGAGAGATGACATGAGG 360
DB 37 LysIleHisArgAlaAlaIleLysGlyAspAlaGluValGluHis---CysLeuThr 55
QY 361 AAGAGCCGCTCAACTGACAAAGAGATATGAAAGAGGACTCTCTACACTGGGCC 420
DB 56 ArgArgPheArgAspLeuAspAlaArgAspArgLysAspArgThrValLeuHisLeuThr 75
QY 421 TGTGTCATGCCATGACANAGTAGTAACATTTCTGGTAGACAGAAAGTCCGCGTTAT 480
DB 76 CysAlaHisGlyArgValGluValValThrLeuLeuLeuSerArgArgCysGlnIleAsn 95
QY 481 GTCTTGTGATCGGAGGAGGACACCTCTGATGAGGCTCTCAATGCCAGAGGGAAGCT 540
DB 96 IleTyrAspArgLeuAsnArgThrProLeuMetLysAlaValHisCysGlnGluAla 115
QY 541 TTGTGCAAAATTTCTCATAGATGCTGGTGTGATCTAAATATTATGTAGATGTGTGCGAA 600
DB 116 -CysAlaIleLeuLeuGluHisGlyAlaAsnProAsnIleLysAspIleTyrSerAs 135
QY 601 CACGGCTCTCCATTATGCCCTTTATAGTGAGAATTTATTAAATGCTGGCAACACTGCTGC 660
DB 135 nThrAlaLeuHisTyrAlaValTyrAsnLysGlyThrSerLeuAlaGluLysLeuLeu 155
QY 661 CTATGTGAGTCAATCAGGTGCAAAACAAAGGTCACGCTCACCCCTTTTACTGGCCAT 720
DB 155 rHisHisAlaAsnIleGluAlaLeuAsnGluGluGlyAsnThrProLeuLeuPheAlaIle 175
QY 721 ACAGAAAGAGCAAGCAAACTGTGGAATTTTACTAACAAAAATGCAAAATGCCAACGC 780
DB 175 eAsnSerArgArgGlnGlnIleValGluPheLeuLysAsnGlnAlaAsnLeuHisAl 195
QY 781 ATTTAATGAGTCTAAATGCACAGCCCTCATGCTTGCCATATGTGAAGGCTCATCAGAGAT 840
DB 195 alleAspAsnPheArgThrAlaLeuMetLeuAlaValGlnHisAsnSerSerIle 215
QY 841 AGTCGGCATGCTCTTCAGCAAAATCTGACGCTCTTGCTGAGACATACATGGAATAC 900
DB 215 eValSerLeuLeuGlnGlnAsnIleAsnIlePheSerGlnAspLeuPheGlyGlnTh 235
QY 901 TGCAGAACGTTATGCTGCTGCTGGAGTAAATACATTTCATCAACACTTTTGGAAACA 960
DB 235 rAlaGluAspTyrAlaValCysTyrAsnPheArgSerIleGlnGlnGlnIleLeuGluHi 255
QY 961 TATACGAAATTCACCTAAAAAT 982
DB 255 sLysAsnLysIleLeuLysSer 262

RESULT 13
Q72468
ID Q72468 PRELIMINARY; PRT; 357 AA.
AC Q72468;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalobso D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Shevchenko Y., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalus D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC056266; AAH56266.1; -
KW Hypothetical protein.
FT NON TER 357 357
SQ SEQUENCE 357 AA; 41216 MW; D6584C8C280ECA65 CRC64;

Alignment Scores:
Pred. No.: 1.95e-21 Length: 357
Score: 499.00 Matches: 127
Percent Similarity: 54.65% Conservative: 55
Best Local Similarity: 38.14% Mismatches: 124
Query Match: 7.74% Indels: 27
DB: 4 Gaps: 6

US-09-602-362E-26 (1-3673) x Q72468 (1-357)

QY 190 CTCTTAGCTCCGCTGCGAAGGGCGTCCGGGGCCCGAGCCCGGACCCCGAACCCCTTCAGCGAA 249
DB 15 LeuLeuSerSerMetAspGlnGluTyrAlaGlyProGly----- 27
QY 250 CGGGTCTACACTGAGAAGGACTACGGGACCATCTACTTCGGGATCTAGGAAAGATCCAT 309
DB 28 -----TyrAspIleArgAspTrp-----GluLeuArgLysIleHis 39
QY 310 ACAGCTGCCCTCCCGGGCCCAAGTCCAGAAAGCTGAGAGATGACATGAGAGAGAGGCC 369
DB 40 ArgAlaAlaIleLysGlyAspAlaAlaGluValGluArg---CysLeuThrArgArgPhe 58
QY 370 GTCAACCTGCAACAAAAGAGATATGAAAGAGAGAGACTGCTCTACACTGGCCCTGTGTCAAT 429
DB 59 ArgAspLeuAspAlaArgAspArgLysAspArgThrValLeuHisLeuAlaCysAlaHis 78
QY 430 GGCCATGCANAAGTAGTAACTTTCTGGGTAGACAGAAAGTGCNGCTTAAATGCTCCTTGT 489
DB 79 GlyArgValGlnValThrLeuLeuLeuHisArgArgCysGlnIleAspIleCysAsp 98
QY 490 GGCGAAGGGAGGACACCTCTGATGAAGGCTCTCAATGCGAGGAGGAACTTTGTGCAAA 549
DB 99 ArgLeuAsnArgThrProLeuMetLysAlaValHisSerGlnGluGluAla-CysAlaIle 118

QY	550	TATTCTCATAGACTGCTGGTCTGATCTAAATATGTAGATGCTGTATGGCAACGCTCT	609
Db	118	eValLeuLeuGluCysGlyAlaAsnProAsnIleGluAspIleYrGlyAsnThrAlaLe	138
QY	610	CCATTATGCCCTTTATAGTGAAGATTTATTAAATGGTGGCAACACTGCTGCTCTATGGTC	669
Db	138	uHsYrAlaValYrAsnLysGlyThrSerLeuAlaGluArgLeuLeuSerHisAl	158
QY	670	AGTCATCAGGTGCAAAACAAGGTAGCTCAGACCCCTTTTACTGCCATACAGAAAG	729
Db	158	aAsnIleGluAlaLeuAsnLysGluGlyAsnThrProLeuLeuPheAlaIleAsnSerAr	178
QY	730	AAGCAAGCAAACTGTGGATTTTACTAACAAAAATGCAAAATGCAAAACGCAITTAATGA	789
Db	178	gArgGlnHisMetValGluPheLeuLeuLysAsnGlnAlaAsnIleHisAlaValAspAs	198
QY	790	CTCTAAATGCCAGCCCTCATCTGCTGCATATGTAAGGCTCATCAGAGATAGTCGGCAT	849
Db	198	nPheLysArgThrAlaLeuIleLeuAlaValGlnHisAsnLeuSerSerIleValThrLe	218
QY	850	GCTTCTTCAGCAAAATGTTGACGCTTTGCTGTAAGACATACATCGAATAACTGCAGAACG	909
Db	218	uLeuLeuGlnGlnAsnIleArgIleSerSerGlnAspMetPheGlyGlnThrAlaGluAs	238
QY	910	TTATGCTGCTGCTGCTGGAGTTAATTACATTCATCAACAACCTTTTGGAAATATACGAA	969
Db	238	pTyRAlaLeuCysSerAspLeuArgSerIleArgGlnGlnIleLeuGluHisLysAsnLy	258
QY	970	ATTACCTAAAATCTCTAAATACCAATCCAGAGGACATCTACAGAACACCTGATGA	1029
Db	258	sMetLeuLysAsnHisLeuArgAsnAspAsnGlnGlnThrAlaAlaMetLysPro----	276
QY	1030	GGCTGCACCTTGGGGAAAGAACCTTGACACGGCTGAAAGCTTGCTGGAAAAAACACC	1089
Db	277	----AlaAsnLeuLysLysArgLysGluArgAlaLysGluHisAsnLeuLysValAlaSe	295
QY	1090	TGACAGAGCTGCACCTGTTGGTGGAGGAAAGCTGCTGCCAAA-----ATTCAATG	1137
Db	295	rGluGluLysGlnGluArgLeuGlnArgSerGluAsnLysGlnProGlnAspSerGlnSe	315
QY	1138	TCGTGGGAAA-----GCAACATCTGGAAGTTT 1165	
Db	315	rTyGlyLysLysLysAspAlaMetTyGlyAsnPhe 327	

RESULT 14

O9HCD1

PRELIMINARY; PRT; 718 AA.

AC O9HCD1,

DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE Hypothetical protein KIAA1641 (Fragment).

GN KIAA1641.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20450683; PubMed=10997877;

RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes.

RT XVIII. The complete sequences of 100 new cDNA clones from brain which

RT code for large proteins in vitro."

RL DNA Res. 7:273-281(2000).

DR EMBL; AB046861; BAB13467.1; -.

KW Hypothetical protein.

FT NON_TER 1

SQ SEQUENCE 718 AA; 82972 MW; 2AD05ADE70E3C97F CRC64;

Alignment Scores: 8.2e-21 Length: 718

Pred. No.: 489.00 Matches: 160

Score:

		Percent Similarity: 42.60%	Conservative: 79
		Best Local Similarity: 28.52%	Mismatches: 186
		Query Match: 7.58%	Indels: 136
		DB:	Gaps: 12
	US-09-602-362E-26 (1-3673) x Q9HCD1 (1-718)		
QY	2009	TCTCTGTTAAAGATGGTCTTCTGAAGCCTACCTGTGGAGGAAGTTTCTCTTCCAAAT	2068
Db	11	SerSerGlnLysGlnProAlaLeuLysAlaThrSerAspLysLysAspSerValSerAsn	30
QY	2069	AAAGCTTAGAATTAAGACAGACAGAACATTCAAAGCAGAGTCTCCTGATAAAGATGCT	2128
Db	31	IleProThrGluIleLysAspGlyGlnGlnSerGlyThrValSerSerGlnLysGlnPro	50
QY	2129	CTTCTGAAGCCTTACCTGTGGAGGAAGTTTCTTCCAAATAAAGCTTGAATTAAG	2188
Db	51	AlaTrpLysAlaThrSerValLysLysAspSerValSerAsnIleAlaThrGluIleLys	70
QY	2189	GACAGAGAAACACTCAAGCAGAGTCTCTGTAATATGATGGTCTTCTGAAGCCTACCTGT	2248
Db	71	AspGlyGlnIleArgGlyThrValSerProGlnLysGlnSerAlaGlnLysValIlePhe	90
QY	2249	GGAGGAAAGTTTCTTCCAAATAAAGCTTTAGAATTAAGGACAGAGAAACATTCAAA	2308
Db	91	LysLysLysValSerLeuAsnIleAlaThrArgIle-----	103
QY	2309	GCAGCTCAGATGTTCCCATCAGATCCAAACAAAGGATGATGAAGAAATTCITGGAT	2368
Db	104	-----ThrGlyLysLys 108	
QY	2369	TTTGAGAGTTTCTTTCAGACTCTCTACAGATGATGTGTGTTTACCAAGGCTACACAT	2428
Db	108	-----	108
QY	2429	CAAAAAGAAATTCGATACCTTAAGTGGAAAATTAAGAGATCTCTCTGATAAAGATGGTCT	2488
Db	109	-----SerGly-----ThrGluTyProGluAsnLeuProThr 119	
QY	2489	CTGAAGCCTTACCTGTGGATGAATTTCTCTTCCAAATAAAGCCTTGAATTAAGGAC	2548
Db	120	LeuLysAlaThrIleGluAsnLysAsnSerValLeuAsnThrAlaThrLysMetLysAsp	139
QY	2549	AGAGAAACATTCAAAGCAGAGGATGTGAGTTCTGTAGATCCACATTCAGTCTTTTGGC	2608
Db	139	-----	139
QY	2609	AAACCGACTACTGAAAATTCACAGTCTCAAAAGTTGAGGAAGCTTTTAACTTACTACC	2668
Db	140	-----ValGlnThrSerThrProGluGlnAspLeuGluMetAlaSer 153	
QY	2669	AAGGGGAGCAACAAAGACAGTAAGTGGACACAGAACGTGATATTGGCATTATTGAA	2728
Db	154	Glu-----GlyGluGlnLysArg-----	159
QY	2729	CGAGCTCCCAAGATCAAAATAAGATGCCACATCAGAAATTAGGAAGAAAAGAGAT	2788
Db	160	-----LeuGluGluTyGluAsnAsnGlnProGlnVal-----LysAsnGln 173	
QY	2789	ACAAAATCAACTTCAGAT---TCTGAGATTATCTGTGATGATACACAGAAATTATGAG	2845
Db	174	IleHisSerArgAspLeuAspIleIleGlnSerSerGlnThrValSer-----	191
QY	2846	TGTTTACCTGAGGCTACATATCAAAAGAAATAAAGACAAACAATGGCAAAATAGAGAG	2905
Db	191	-----	191
QY	2906	TCTCTGAAAAGCCTTCTCACTTTTGAGCCTGCCACTGAAATGCAAACTCTGTTCCTCAAT	2965
Db	192	-----GluAspGlyAspSerLeuLysCysCysAsnCysLysAsn 203	
QY	2966	AAAGGCTTAGAATGAAGAATAAACACATTTGAGAGCAGATTCAACTACCTTATCAAA	3025
		-----	---

Db 417 laGluValGluSerLeu-----HisSerSerLeuAlaThrAlaIleAsnGluTyrA 434
QY 1797 AAAAGACGCTTCCAAATTAAGCCTTTGAATTGAAGATGAACAACATTGAGAGAGCTC 1856
Db : : : : :
Db 434 snGluIleValGluArgLysAspLeuGluLeu-----ValLeuTrpArgAla- 449
QY 1857 AGATGTTCCCATCAGAAATCCAAACAAAGAGCAGTGAAGAAAATTTCTGGGATTCTGAGA 1916
Db : : : : :
Db 450 -----AspAspValSerArgHisGluLysMetGlyS 460
QY 1917 GTCCCTGTGAGACGGTTTCACAGAGGATGTGTTATTTACCAAGCTTACACATCAAAA- 1975
Db : : : : :
Db 460 erAsnIleSerGlnLeuThrAspLysAsnGluLeuLeuThrGlnGluValHisLysAlaA 480
QY 1976 -----GAATTCGATACCTTAAAGTGAAGAAATTTAGAGAGCTCCCTGTTAAAGATGCTCTTC 2030
Db : : : : :
Db 480 rgValLysPheAsnThrLeuLysGlyLysLeuArgGluThr-----ArgAspAlaLeuA 498
QY 2031 TGAAGCCTACCTGTGGAGGAAAGTTTCTTCCAAATAAAGCCTTACAATTAAAGGACA 2090
Db : : : : :
Db 498 rgGluLysThr-----LeuAlaLeuGlySerValGlnLeuAspLeuArgGlnA 514
QY 2091 GAGAA-----ACATTCAAAGCAGAGTCTCCTGTATAAGATGCTCTTCTGAAGC 2138
Db : : : : :
Db 514 laGlnHisArgIleLysGluMetLysGlnMetHisProAsnGlyGluAlaLysGluSerG 534
QY 2139 CTACCTGTGGAAGAAAGTTTCTTCCAAATAAAGCCTTAGAATTAAAGCAGAGAAA 2198
Db : : : : :
Db 534 lnSerIleGlyLysGlnAsnSerLeuGluGluArgIle----- 546
QY 2199 CACTCAAGCAGAGTCTCCTGATATGATGGTCTCTGAAGCCTACCTGTGGAGGAAAG 2258
Db : : : : :
Db 547 -----ArgGlnG 549
QY 2259 TTTCTCTTCCAAATAAAGCTTTAGAATTGAAGGACAGAGAAACATTCAAAGCAGCTCAGA 2318
Db : : : : :
Db 549 lnGluLeuGluAsnLeuLeuGlu----- 557
QY 2319 TGTTCCCATCGAATCCAAACAAAGGATGATGAAGAAATTTCTGGGATTTGAGAGTT 2378
Db : : : : :
Db 558 -----ArgGlnLeuGluAspAlaArgLysGluGlyAspAsnLysGluI 572
QY 2379 TCCTTGAGACTCTTACAGATGATGTGTTTACCAGAGCTTACACATCAAAAGAAAT 2436
Db : : : : :
Db 572 leValIleAsnIleHisArgAsp-----CysLeuGlu----- 582
QY 2439 TCGATACCTTAAGTGGAAAA-----TTAGAAGTCTCCTGATAAGATGGTCTTCTGA 2492
Db : : : : :
Db 583 -----AsnGlyLysGluAspLeuLeuGluGluArgAsnLysGlu-----LeuMetL 598
QY 2493 AGCTACCTGTGGAAATTTCTCTTCCAAATAAAGCCTTAGAATTGAAGGACAGAG 2552
Db : : : : :
Db 598 ysGluTyrAsnTyrLeuLysGluLysLeuGlnCysGluLysGluLysAlaGluArgG 618
QY 2553 AAACA-----TTCAAAGCAGAGGATGTGAGTCTCTGAGAGTCCACATTCAGTC 2600
Db : : : : :
Db 618 luValIleValArgGluPheGlnGluLeuValAspHisLeuLys-----ThrPheSerI 637
QY 2601 TTTTGGCAACCGACTACTGAAATTCACAGTCTACAAAAGTTGAGGAGACTTTAATC 2660
Db : : : : :
Db 637 leSerGluSerProLeuGluGlyThrSerHisCys-----HisIleAsnL 652
QY 2661 TTACTACAGGAGGAGGACCAAGACAGACTAAGTGAACAGGACGTTGATTTGGCA 2720
Db : : : : :
Db 652 euAsnGlu-----ThrTrpThrSerLysLysLeuPheG 664
QY 2721 TTATTGACGAGTCCCAAGATCAACAATAAGATGCCACATCAGAATTAGGAAGAA 2780
Db : : : : :
Db 664 lnValGluIleGlnProGluGlu-----LysH 673
QY 2781 AAGAAGATCAAAATCAACTTCAGATTCCTGAGATTATCTCTGTGAGTGATACACAAAT 2840
Db : : : : :
Db 673 isGluGluPheArgLysLeu-----PheGlnLeuIleSerLeu-----LeuAsnT 688

RESULT 2

POTE HUMAN

ID POTE HUMAN STANDARD; PRT; 584 AA.
AC Q86V6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE POTE protein (Prostate, ovary, testis expressed protein).

QY 2841 ATGAGTGTTCCTGAGGCTACATATCAAAAAGAAATAAGACAAACAATGCGAAATAG 2900
Db : : : : :
Db 688 yrThrAla-----AspGlnIleArgLysLysAsnArgLysLeuG 701
QY 2901 AAGAGTCTCTGTAAGAGCCTTCTCAC-----TTTGAGCCTGCGACCTGAATCAAA 2951
Db : : : : :
Db 701 luGlu-----GluAlaThrGlyTyrLysLysCysLeuGluMetThrIleAsnMetLeuA 719
QY 2952 ACTCTGTTCCTCAATAAAGGCTTAGAATGGAAG-----AATAAACAAACATTGAGAG 3002
Db : : : : :
Db 719 snAlaPheAlaAsnGluAspPheAsnCysHisGlyAspLeuAsnThrAspGlnLeuLysM 739
QY 3003 CAGATTCAACTACCTATCAAAAATCTTGATGCACCTTCTCTTCTGTGAAAGGAGGAGG 3062
Db : : : : :
Db 739 etAsp---lleLeuPheLysLysLeuLysGlnLysPheAsnAspLeuValAlaGluLysG 758
QY 3063 AACTTTAAAAAGATACTGTGAACAAATACAGCAAAAAATGGAACAAATCAAAAAATAAGT 3122
Db : : : : :
Db 758 luAlaValSerSerGluCysValAsnLeuAlaLysAspAsnGlu----- 772
QY 3123 TTTGTGTACTACAAAGGAACTGTGAGAAGCGAAGAAATAAAA-----TCACAGT 3173
Db : : : : :
Db 773 -----ValLeuHisGlnGluLeuLeuSerMetArgAsnValGlnGluLysCysGluLysL 791
QY 3174 TAGAGAACCAAAAAGCTAAATCGGAACAAGAGCTCTGCACTGTGAGA----- 3220
Db : : : : :
Db 791 euGluLysAspLysLysMetLeuGluGluValLeuAsnLeuLysThrHisMetGluL 811
QY 3221 -----TTGGCTTTAAATCAAGAGAAAGAGAGAGAAATGTCGAT-----ATAT 3266
Db : : : : :
Db 811 ysAspMetValGluLeuGlyLysLeuGlnGluTyrLysSerGluLeuAspGluArgAlaV 831
QY 3267 TAAAGAAAAAATTAGCCCGAAGAG-----CAACTTAGGAAGAAGTTAGAGTGAACACC 3323
Db : : : : :
Db 831 aGlnGluIleGluLysLeuGluGluIleHisLeuGlnLysGlnAlaGluTyrGluLysG 851
QY 3324 AACTTGAACAG-----ACTCTCAGATACACAAGATATAGAAATGGA 3362
Db : : : : :
Db 851 lnLeuGluGlnLeuAsnLysAspAsnThrAlaSerLeuLysLysLysGluLeuThrLeuL 871
QY 3363 AAGGTCTACAAAGTAATTGATCAGGTTTCTCACACTCATGAA-----AGTGAAA 3413
Db : : : : :
Db 871 ysAspValGluCysLysPheSerLysMetLysThrAlaTyrGluValThrGluL 891
QY 3414 ATGATCTCTTTCATGAA-----AATGTCATGTTGAAA- 3445
Db : : : : :
Db 891 euGluGluPheLysGluValPheAlaAlaValLysAlaAsnAsnSerMetSerLysL 911
QY 3446 -----AAGGAATTGCCATGCTTAAACTGGAAGTACGCCACACTGAAAC 3488
Db : : : : :
Db 911 ysLeuMetLysSerAspLysLysIleAlaValIleSerThrLysLeuPheThrGluLys- 930
QY 3489 ATCAACACCCAGGTGAAGGAAAAATAAATACTTT-----G 3521
Db : : : : :
Db 931 -----GlnArgMetLysTyrPheLeuSerThrLeuProThrArgProGluP 946
QY 3522 AGGACATTAAAGATTTTACAAGAAAAGATGCTGAATCTCAAAATGACCTCAAACTGAAAC 3581
Db : : : : :
Db 946 roGluLeuProCysValGluAsnLeuAsn-----SerIleGluLeuAsnA 961
QY 3582 AGAACAACAGTAAACAAAAGGCA 3604
Db : : : : :
Db 961 rGlyTyrIleProLysThrAla 968

GN POTE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RC TISSUE=Prostate;
RX MEDLINE=2238248; PubMed=12475935;
RA Bera T.K., Zimonjic D.B., Popescu N.C., Sathyanarayana B.K., Kumar V.,
RA Lee B., Pastan I.;
RT "POTE, a highly homologous gene family located on numerous
RT chromosomes and expressed in prostate, ovary, testis, placenta, and
RT prostate cancer"; Proc. Natl. Acad. Sci. U.S.A. 99:16975-16980(2002).
RL [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Kump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.B.,
RA Minoshima S., Shimizu N., Nagasiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21";
RT Nature 405:311-319(2000).
CC -!- SUBCELLULAR LOCATION: Basal and terminal prostate epithelial
CC cells.
CC -!- TISSUE SPECIFICITY: Expressed in prostate, ovary, testis, placenta
CC and prostate cancer cell lines.
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AY172978; AAC03914.1; -;
CC DR EMBL; AP001465; -; NOT_ANNOTATED_CDS.
CC DR XIM; 607549; -;
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; ank; 7.
CC SMART; SM00248; ANK; 6.
CC PROSITE; PS00088; ANK REPEAT; 5.
CC PROSITE; PS0297; ANK REP REGION; 1.
CC Repeat; ANK repeat; Polymorphism.
FT REPEAT 172 201 ANK 1.
FT REPEAT 205 234 ANK 2.
FT REPEAT 238 267 ANK 3.
FT REPEAT 271 300 ANK 4.
FT REPEAT 304 333 ANK 5.
FT REPEAT 337 366 ANK 6.
FT VARIANT 113 113 S -> G (in dbSNP:65178691).
FT VARIANT 135 135 V -> I (in dbSNP:65178701).
FT FTID=VAR_016242.
FT FTID=VAR_016243.
SQ SEQUENCE 584 AA; 66378 MW; CB0199694EC63340 CRC64;

Alignment Scores:
Pred. No.: 4,87e-22 Length: 584
Score: 588.00 Matches: 176
Percent Similarity: 46.36% Conservative: 98
Best Local Similarity: 29.78% Mismatches: 208
Query Match: 9.12% Indels: 109

DB: 1 Gaps: 16
US-09-602-362E-26 (1-3673) x POTE_HUMAN (1-584)
QY 124 GGGCTGGGAGGCGGAGGCGGCGGCTCTCTAGCAGGCGGCTGCAGCCATG 183
Dy 74 GlySerGlyThrSerAsnValGlyThrSerGlyAsp-----HisGluAsnSerPheMet 91
QY 184 AAGAGGCTCTTAGTGTCCGCTGGCAAGGCG-----GTCCGGGGC 222
Dy 92 LysMetLeuArgSerLysMetGlyLysTrpCysHisCysPheProCysCysArgGly 111
QY 223 CCGGAGCCCGCAAC-----CCCTTCAGCGAAGCGGTCTACACTGAGAAG 267
Dy 112 SerSerLysSerAsnValGlyAlaTrpGlyAspTrpAspHisSerAlaPheMetGluPro 131
QY 268 GACTACGGGACCATCTACTTCGGGGATCTAGGAGGATCCATACAGCTGCTCCCGGGC 327
Dy 132 ArgTyr---HisValArgGluAspLysLeuHisArgAlaAlaTrpTrpGly 150
QY 328 CAAGTCCAGAGCTGGAGAGATGACAGTAGGAGAGAGCCGTCACCTGAACAAAGA 387
Dy 151 LysValProArg---LysAspLeuValMetLeuArgAspThrAspMetAsnLysArg 169
QY 388 GATATGAAGAGAGGACTGCTCTACACTGGGCGCTGTGTCATGGCCATGCANAAAGTAGTA 447
Dy 170 AspLysGluLysArgThrAlaLeuHisLeuAlaSerAlaAsnGlyAsnSerGluValVal 189
QY 448 ACATTTCGTAGACAGAAAGTCGCGCTTAATGTCCTGATGGGAGGAGGAGACCT 507
Dy 190 GlnLeuLeuAspArgArgCysGlnLeuAsnValLeuAspAsnLysLysLeuAla 209
QY 508 CTGATGAAGCTCTACATGCGAGGAGAGCTTTGTGCAATATCTCATAGATGCTGG 567
Dy 210 LeuLeuLysAlaLeuGlnCysGlnGluAspGlu-CysValLeuMetLeuLeuGluHisG 229
QY 568 TGTGATCTAAATATTATGTATGGCAACAGCGCTCTCCATTATGCGGTTTATAG 627
Dy 229 YalaAspArgAsnLeuProAspGluTy-GlyAsnThrAlaLeuHisTyAlaLeuTyrAs 249
QY 628 TGAGAAATTATTAAATGGGCAACCTGCTGCTCTGTCAGTCAGTCAGGTCAGGTCAGG 687
Dy 249 nGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyGlyAlaAspLysGluSerLysAs 269
QY 688 CAAGGTGCTGCTCACACCCCTTTTACTGGCCATACAGAAAGAGAGCAAGCAACAGTGTGA 747
Dy 269 nLysCysGlyLeuThrProLeuLeuLeuGlyValHisGluGlnLysGlnValVal 289
QY 748 ATTTTCTAACAAATAATGCAACCGCATTTAATGAGTCTAAATGCGACAGCCCT 807
Dy 289 sPheLeuLeuLysLysAlaAsnLeuValLeuAspArgTyrGlyArgThrAlaLeu 309
QY 808 CATGCTTGCATATGTAAGGCTCATCAGAGATAGTCGGCATGCTTCTTCAGCAAAATGT 867
Dy 309 uilleLeuAlaValCysGlySerAlaSerileValaenLeuLeuLeuGluAsnVa 329
QY 868 TGAAGCTTTTGTGCAAGACATACATGAATACTGAGAACCGTTATGCTGCTGCTGCTGG 927
Dy 329 lAspValSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyAlaValSerSerHi 349
QY 928 AGTTAATTACATTCACCAACTTTTGAACATATACGAAAA-----TTACCTAAAAA 981
Dy 349 sHisHisValilleCysGluLeuLeuSerAspTyLysGluLysGlnMetLeuLysLe 369
QY 982 TCCTCAAAATACCAATCCAGAGAGCAATCTACAGGAACACCTGATGAGGCTGCACCTT 1041
Dy 369 rSerGluAsnSerAsnProGlu----- 376
QY 1042 GCGGAAAGNACACCTGACAGCGCTGAAAGCTTGTGGAAGAAAAACACCTGACGAGCGCTG 1101
Dy 377 -----GlnAspLeuLysLeuThrSerGluGluSerGlu 388
QY 1102 ACGCTTGTGGAGGAGACGCTCTGCCAAAATTCATGCTCTGGGAAAGCAACATCTCGAAA 1161

Db 388 mArgLeu----- 390
 QY 1162 GTTTGAACAGCTCAACAGAGAAACACCTAGGAAATTTTGGAGGCTTACAAAAGAACATC 1221
 Db 391 ----LysValSerGluAsnSerGlnProGluLysMet----- 401
 QY 1222 TGAGAAATTTTCATGTGGCAGCAAAAGAAAGATCTAGGAATCATCATGGAGGAGAAAGA 1281
 Db 402 -----SerGlnG 404
 QY 1282 AACATCTGTAAGACTGAATGCTGCGAGGAGTAACACCTAATAAACTGAAGTTTGA 1341
 Db 404 uProGluLeuAsnLysAspCysAspArgGluVal-----GluGluGluLeu 420
 QY 1342 AAAAGGAAACATCTAATATGATTCATGCTCTACAAAAGAAACATCTACAAAAGCAAGTAC 1401
 Db 420 sLysHisGlySerAsnProValGlyLeuProGluAsnLeuThrAsnGlyAlaSerAlaG 440
 QY 1402 AAATGTGATCTAGTCTCTAGAGCTTATATTCAGTCTTTTGGACACGGACTATTGA 1461
 Db 440 YAsnGlyAsp-----AspGlyLeuLeuProGlnArgSerArgLysProG 456
 QY 1462 AAATTCACAGTGTACAAAAGTTGAG-----CAAGACTTTTAACTCTGTC 1503
 Db 456 uAsnGlnGlnPheProAspThrGluAsnGluGluTyHisSerAspGluGlnAsnAsp 476
 QY 1504 TACCAAGATATCTCTAAGAGTCTGCACAGAAATATATACGTGTTTACCTGAT----- 1555
 Db 476 rArgLysGlnLeuSerGluGlu-----GlnAsnThrGlyLeuSerGlnAspGluLeu 494
 QY 1556 -GCTACATATCAAAAGATATCAAAACATAAATACAAATAGATAGATCAGATGTTCC 1614
 Db 494 uThrAsnLysGlnLysGlnLeuGluValAlaGluGlnLysMetAsnSerGluLeuSerLe 514
 QY 1615 ATCAATATCCAAACGACGAGGAGATGAAGAATATCTTGGATCTGGAGTCTCTTGA 1674
 Db 514 uSerHisLysLys-----GluGluAspLeuLeuArgGluAsnSerValLeuGlnG 531
 QY 1675 GAGTCTCGAAAGACTCAAGTGTATACCTGATCTATCTATCATCAGAAAGTATGAGAT 1734
 Db 531 uGluLeuAlaValLeuArgLeuGluLeuAspGluThrLysHisGlnAsnGlnLeuArg 551
 QY 1735 AAATAGA-----GAAGTAGAAGAGCTTCCTCGAAGCCATCTCGCTTCAAGCCTGC 1785
 Db 551 uAsnLysIleLeuGluGluLeuGluSerValLysGluLysThrAspLysLeuLeuArg 571
 QY 1786 CGTNGAATCCAAAGACTGTTTCCAAATAAA 1816
 Db 571 aMetGlnLeuAsnGluGluAlaLeuThrLys 581

RESULT 3

ANR7_HUMAN STANDARD; PRT; 201 AA.
 AC Q2527; Q96QN1; Q9UDM3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ankyrin repeat domain protein 7 (Testis-specific protein TSA806).
 GN ANKRD7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Testis;
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey M., Helton B., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE OF 51-201 FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=96411689; PubMed=8812458;
 RA Ozaki K., Kuroki T., Hayashi S., Nakamura Y.;
 RT "Isolation of three testis-specific genes (TSA303, TSA806, TSA903) by
 a differential mRNA display method.";
 RL Genomics 36:316-319 (1996).
 RN [3]
 RP SEQUENCE OF 8-138 FROM N.A.
 RA Courtney L., Weidmann P., Johnson D., Tucci S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- TISSUE SPECIFICITY: Testis specific.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; BC032799; AAH32799.1; -;
 DR EMBL; BC007020; AAH07020.1; -;
 DR EMBL; D78334; BAA11348.1; -;
 DR EMBL; AC007874; AAF19231.1; -;
 DR Genew; HGNC:18588; ANKRD7.
 DR HSP; P25963; I1KN.
 DR GO; GO:0008584; P:male gonad development; TAS.
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00023; ank; 5.
 DR SMART; SM00248; ANK; 5.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 KW ANK repeat; Repeat.
 FT REPEAT 5 34 ANK 1.
 FT REPEAT 38 67 ANK 2.
 FT REPEAT 71 100 ANK 3.
 FT REPEAT 104 133 ANK 4.
 FT REPEAT 137 166 ANK 5.
 FT CONFLICT 51 55 NEDCA -> MBIVP (IN REF. 2).
 SQ SEQUENCE 201 AA; 22588 MW; 30C241795D2C049F CRC64;

Alignment Scores:

Pred. No.: 5.98e-11 Length: 201
 Score: 362.00 Matches: 77
 Percent Similarity: 65.68% Conservative: 34
 Best Local Similarity: 45.56% Mismatches: 57
 Query Match: 5.61% Indels: 1
 Ds: 1 Gaps: 0

US-09-602-362E-26 (1-3673) x ANR7_HUMAN (1-201)

QY 385 AGAGATATGAAGAAGAGAGACTGCTCTACCTGGCGCTGTGTCAATGGCCATGCATGCAAGTA 444

Db 2 GlnAspLysLysTyrArgThrProLeuHisLeuAlaCysAlaAsnGlyHisThrAspVal 21

Db	790	uLysGlnIleLysHisLeuGluIleGluLysAsnAlaGluSerSerLysAlaSerSerIl	810	Db	1110	nGluLeuAspThrLeuLysGluAsnValGluGluLeuAsnLysSerLy	1130
Qy	2233	TCTGAGCCTACCTGTCGAGGAAAGTTCTCTCCAAATAAAGACTTTAGAAATGAAGGA	2292	Qy	3193	ATGGGAACAAGAGCTCTCCAGTGTGAGATTCCTTAAATCAAGAA	3249
Db	810	eThrArgGluLeuGlnGlyArgGluLeuLysLeuThrAsnLeuGlnGluAsnLeuSerGI	830	Db	1130	s-----GluLeuLeuThrValGlu-----AsnGlnLysMetGluGluPheAr	1144
Qy	2293	C-----ACAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACA	2340	Qy	3250	AAGAAATTCGATATATATATAA-----	3271
Db	830	uValSerGlnValLysGluThrLeuGluLysGluLeuGlnIleLeuLysGluLysPheAl	850	Db	1144	gLySGluIleGluThrLeuLysGlnAlaAlaGlnLysSerGlnGlnLeuSerAlaLe	1164
Qy	2341	AAAGATCATCAAGAAATCTCTGGGATTTTGAGAGTTTCTTGAGACTCTCTTACAGAA	2400	Qy	3272	-----GAAATAATTAGACCCGAGCAGCACTTAGGAAAG-----	3307
Db	850	aGluAlaSerGluGluAlaValSerValGlnArgSerMetGlnGluThrVal-----	867	Db	1164	uGlnGluGluAsnValLysLeuAlaGluGluLeuGlyArgSerArgAspGluValThrSe	1184
Qy	2401	TCATGTGTGTTTACCAAGGCTACACATCAAAAGAA-----TTCGATACCTTAAAGTGG	2454	Qy	3308	-----	3321
Db	868	-----AsnLysLeuHisGlnLysGluGluGlnPheAsnMetLeuSerSe	882	Db	1184	rHisGlnLysLeuGluGluGluArgSerValLeuAsnAsnGlnLeuLeuGluMetLysLy	1204
Qy	2455	AAATTAGAAGTCTCTGTAAGATGTCCTTCTGAAGCCTACCTGTGGAATGAAAT	2514	Qy	3322	CCAACTTGAACAGACTCTCAGAATACAGATATAGAAATGAAAGTGTACAAAGTAATTT	3381
Db	882	rAspLeuGluLys-----LeuArgGluAsnLeuAlaAspMetGluAl	896	Db	1204	sArgGluSerLysPheIleLysAspAlaAspGluGluLysAlaSerLeuGlnLysSerIl	1224
Qy	2515	TTCTCTTCCAAATAAAGCCTTAGAATTGAAGCAGAGAGAAACATTCAAAGCAGAGGATGT	2574	Qy	3382	GAATCAGGTTTCTCACACTCATGAAAGTGAATATGATCTCTTCATGAAATTCGATGTT	3441
Db	896	alysPheArgGluLys-----AspGluArgGluGluGlnLeuIleLysAlaLysGlu--	913	Db	1224	eSer-----IleThrSerAlaLeuLeuThrGluLysAspAlaGluLeuGluLys-----Le	1241
Qy	2575	GAGTCTCTGAGTCCACATTCAGTCTTTTGGCAACCCAGTACTGAAAT-----TCACA	2631	Qy	3442	GAAAGGAAATTCGCATGCTAAATCTGAAAGTACCCACACTGAAACATCAACACAGGT	3501
Db	914	----LysLeuGluAsnAspIleAlaGluIleMetLysMetSerGlyAspAsnSerGI	932	Db	1241	uArgAsnGluValThrValLeuArgGlyGluAsnAlaSerAlaLysSerLeuHisSerVa	1261
Qy	2632	GTCACAAAGTCTGAGGAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGT	2691	Qy	3502	GAAAGGAAATTAATCTTTGAGGACATTAAGATT-----TTACAAGAAAGAAATGCTGAAC	3558
Db	932	nLeuThrLysMetAsnAspGluLeuArgLeu-----	942	Db	1261	lValGlnThrLeuGluSerAspLysValLysLeuGluLysValLysAsnLeuGluLe	1281
Qy	2692	AACTGGCAACAGGACCTGTATTT-----	2716	Qy	3559	TCAATGACCTTAAATCAACACAG	3583
Db	943	-----LysGluArgAspValGluGluLeuGlnLysLeuThrLysAlaAsnGI	959	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	2717	-----GGCATTATTGACAGCTCCACAGATCAACAAATTAAGATGCCCATCA--	2767	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	959	uAsnAlaSerPheLeuGlnLysSerIleGluAspMetThrValLysAlaGluGlnSerGI	979	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	2768	-----GAATTAGGAGAAAGAAAGATACAAATCAACTCAGA	2805	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	979	nGlnGluAlaLysLysHisGluGluLysLysGluLeuGluArgLysLeuSerAs	999	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	2806	TTCTGAGATATCTCTGTGAGTGATACAGAAATATGAGTTTACCT---GAGGCTAC	2862	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	999	pleuGlu-----LysLysMetGluThrSerHisAsnGlnCysGlnGluLeuLysAlaAr	1017	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	2863	ATATCAAAAA-----GAAATAAGACAAACAAATGGCAAAATAGAGAGTCTCTCTCA	2913	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	1017	gTyrgluArgAlaThrSerGluThrLysThrLysHisGluGluIleLeuGlnAsnLeuGI	1037	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	2914	AAAG-----CCTTCTCACTTTGAGCTCCACTGCAATGCAAACTCTGT	2958	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	1037	nllyThrLeuLeuAspThrGluAspLysLysGlyAlaArgGlu---GluAsnSerGI	1056	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	2959	TCCAAATAAGGCTTAGATGGAACAAATTAACAAATTCAGAGCAGATTCACCTCCCT	3018	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	1056	yLeuLeuGlnGluLeuGluLeuLeuArgLysGlnAlaAspLysAlaLysAlaGlnTh	1076	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	3019	ATCAAAATCTTGTGATGCTCTCTCTTGTGAAAGAGAGGAACTTAAATAAGATAA	3078	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	1076	rAlaGlu-----AspAlaMetGlnIleMetGlu-----GlnMetThrLysGluLy	1091	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	3079	CTGTGAACAATACAGCAAAATGGAACAATGAAATTAAGTTTGTGTACTACAAA	3138	Db	1281	uGlnLysGluAsnLysArgGln	1289
Db	1091	sThrGluThrLeu---AlaSerLeuGluAspThrLysGlnThrAsnAlaLysLeuGlnAs	1110	Db	1281	uGlnLysGluAsnLysArgGln	1289
Qy	3139	GGAACCTGTCAAGAGCAAGAA-----ATAAATCAGAGTTAGAGAACCAAAAGCTAA	3192	Db	1281	uGlnLysGluAsnLysArgGln	1289

RESULT 5
USOL YEAST
ID USOL YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INTR OR YDLO58w.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RC Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RX Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RC Bai Y., Symington L.S.;
RX Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Required for protein transport from the ER to the Golgi
CC complex.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE

940
1723 AGTAATGGAGATAAATAGAGAAGTAGAAGAGCTTCTCTGAGAAGCCATCTGCCTTCAAG-- 1780
941 --:::|||||:::--LysGluLeuValGluTyrLysSerArgPheGlnSe 952
1781 -----CCTGCCGTNGAAATGCAAAAGACTGTTCCAAAT-----AA 1815
952 rHisAspAsnLeuValAlaLysLeuThrGluLysLeuLysSerLeuAlaAsnTyrIly 972
1816 AGCCTTTGAATTGAAGAAATGAACAAACATTGACAGCAGCTCAGATGTTCCATCAGAACT 1875
972 sAspMetGlnAlaGluAsnGluSerLeuIleLysAlaValGluLysSerLysAsnGluSe 992
1876 CAAA---CAAAGGACGATGAGAAAATCTTGGGATCTGAGAGTCCCTGTCGACAGCT 1932
992 rSerIleGlnLeuSerAsnLeuGlnAsrLysIleAspSer-----Me 1006
1933 TTCACAGAAGGATGTGTATTTTACCCAAGCTACACATCAAAAAGAAATTCGATACCTTAAG 1992
1006 tSerGlnGlu-----LysGluAsnPheGlnIleGluArgGlySerIleGI 1021
1993 TGGAAAATTAGAGAGTCTCTGTGTTAAAGATGCTTCTCGAAGCCTACCTGTGGAAGGAA 2052
1021 uLysAsnIleGluGln-----LeuLysLysThr----- 1030
2053 AGTTTCTCTTCCAATAAAGCCTTAGAATTAAGGACAGAGAACAATTCAAAGCAGAGTC 2112
1031 -----IleSerAspLeuGluGlnThrLysGluGluI 1041
2113 TCCTGATAAAGATGCTCTCTGAAACCTCACTCTGGAAGGAAAGATTCTCTTCCAAATAA 2172
1041 eIleSerLysSerAspSerLysAspGluTyrGluSerGlnIleSerLeuLysGluGI 1061
2173 AGCCTTAGAA-----TT 2184
1061 uLysLeuGluThrAlaThrThrAlaAsnAspGluAsnValAsnLysIleSerGluLeuTh 1081
2185 AAAGACACAGAAAACATCTCAAGCAGAGTCTCTCTGATAATGATGGTCTTCTGAAGCCTAC 2244
1081 rLysThrArgGluGluLeuGluAlaGluLeuAlaLysAsn--LeuLysAsnGI 1100
2245 CTGTGGAAGGAAAGTTCTCTTCCAAATAAGCTTTA---GAATTGAAGACACAGAGAAAC 2301
1100 uLeuGluThrLysLeuGluThrSerGluLysAlaLeuLysGluValLysGluAsnGluGI 1120
2302 ATTCAAAGCAGCTCAGATGTTTCCCATCAGAAATCCAAACAAAGGATGATCAAGAAATTC 2361
1120 uHisLeuLysGluGluLysIleGlnLeuGluLysGluAlaThrGluThrLysGlnGlnLe 1140
2362 TTGGGATTTTCAGATTTCTTCAGACTCTCTTACAGAAATGATGTGTGTTTACCAGAGGC 2421
1140 uAsnSerLeuArgAlaAsnLeuGluSerLeu----- 1150
2422 TACACATCAAAAGAAATTCGATCACTCTTAAGTGGAAATATTAGAAGTCTCCTGTATAAGA 2481
1151 -----GluLysGluHisGluAspLeuAlaGlnLeuLysLysTyrGluGlu----- 1166
2482 TGGTCTTCTGAAGCCTACCTCTGGAATGAAATTTCTCTTCCAATAAAGCCTTAGAATT 2541
1167 -----GlnI 1168
2542 GAAGACACAGAAAACATTCAAGACAGGATGTGAGTTCTGTAGAGTCCACATTCAGTCT 2601
1168 eAlaAsnLysGluArgGlnTyrAsnGluGluIleSerGlnLeuAsnAspGluIleThr-- 1187
2602 TTTTGGCAAAACCGACTACTGAAAATTTACAGTCTACAAAAGCTTCAGGAAGACTTTAATCT 2661
1188 -----SerThrGlnGlnGluAsnGluSerIleLysLysLysAsnAsp----- 1201
2662 TACTACCAAGGAGGAGCAACAAACAGATGAACCTGGACACAGAGAACTGATATTGGCAT 2721
1202 ---GluLeuGluGlyGluValLysAlaMetLysSerThrSerGluGluGlnSerAsn-- 1219

QY	2722	TATTGAACGAGCTCCACAGATCAAAACAATAAGATGCCACATCAGAATTATGAAGAAA	2781
Db	1220	-LeuLYLysSerGluIleAspAlaLeuasn---LeuGlnIleLySGluLeuLYSLy	1238
QY	2782	AGAAGATACAAAAATCAACTTCA-----GATTCTCAGATTAT	2817
Db	1238	sASnGLuThrAsnGLuAlaSerLeuLeuGluSerIleLYSSerValGluSerGluThrVa	1258
QY	2818	CTCTGTGAGTGATACACAGAATTATGAGTGTTTCACCTGAGGCTACATATCAAAGAAAT	2877
Db	1258	IlySIleLeuGLuLeuGlnasp--Glucys-----AsnPhelySGluLYSGLuVa	1274
QY	2878	AAAGACAACAATGGCAAAATAGAAGAGTCTCCTGAAAGCCTTCTCATCTTTGAGCGCTGC	2937
Db	1274	IserGluLeuGLuasplysLYSLySaIaSerGluAAspLYSAsnSerLYSTyr-----	1291
QY	2938	CACTGAATTCAAAACACTCTGTTCCAATAAAGCTTAGAATGGAAGAATAAACAAACATT	2997
Db	1292	-LeuGluLeuGlnLYSGlu-----SerGluLYSIleLYSGluGluLe	1305
QY	2998	GAGAGCAGATTCAACTACCCTATCAAAAATCTTTGGATGCATTCCTTCTTGTGAAAGAGG	3057
Db	1305	uaSPAlaLYSThrThrGluLeuLYSIleGinLeuGluLYSIleThrAsnLeuSerLYSaI	1325
QY	3058	AAGGGAA-----CTTA AAAAGATAACTGTGAACAA-----	3088
Db	1325	aLYSGluLYSerGLuSerGLuLeuSerArgLeuLYSThrSerGLuGluArgLY	1345
QY	3089	-----ATTACAGCAAAATGCAACAAATGAAAAATAAGTTTTGTGTA-----	3130
Db	1345	sASnaIaGluGlnLeuGluLYSLyLeuLYSAsgnGluIleGlnIleLYSAsgnGlnAlaPh	1365
QY	3131	-----CTCAAAAGCAACTGTGCAGA	3150
Db	1365	eGluLYGLuarGLySLeuleuAsnGluGLySerSerThrIleThrGlnGluTYrSerGI	1385
QY	3151	A-----GC	3153
Db	1385	uLYSIleAsnThrLeuGluAspGluLeuIleaArgLeuGlnAsnGluLeuLYSaI	1405
QY	3154	GAAGAAAATA-----AAATCACAGTTAGAG-----	3178
Db	1405	aLYSGluIleAspAsnThrArgSerGluLeuGluLYSVaISerLeuSerAsnAspGluLe	1425
QY	3179	-----AACC AAAAGCTAAATGGGAACAAGAGCTCTGCAGTGTGCAGATT	3222
Db	1425	uLeuGluGluLYSGlnAsnThrIleLYSSerLeuGlnAspGluIleLeuSerTYrLYSaS	1445
QY	3223	GCCTTTAAATCAAGAAGAAGAGAG-----ACAAGAANTGT	3258
Db	1445	pLYSIleThrArgAsnAspGluLYSLeuLeuSerIleGluArgAspAsnLYSArgSpLe	1465
QY	3259	CGATATATTAACAAAAAATTAGACCCGAAGAG-----	3292
Db	1465	uGLuSerLeuLYSGluGlnLeuArgAlaIalaGlnGLuSerLYSaIalYSvalGluGluGI	1485
QY	3293	-----CAACTTAGGAAAAAGTTAGA	3312
Db	1485	yLeuLYSLySLeuGluGluGLuSerSerLYSGluLYSaIaGluLeuGluLYSerLYSGLI	1505
QY	3313	AGTGAACACCAACTTGAACAGACTCTCAGNATACAGATATAGAATTGAAAGCTGTAAAC	3372
Db	1505	uMetMetySLySLeuGluSerThrIleGluSerAsnGluThrGluLeuLYSSerSerMe	1525
QY	3373	AAGTAATTTGAATCAGGTTTCTCAC-----ACTCATGAAG	3408
Db	1525	tGLuThrIleargLYSSerAspGluLYSLeuGluGlnSerLYSSerAlaGluGuas	1545
QY	3409	TGAANAATGATCTTTTCATGAANAATTCGATGTTGAAAAAGAAATGCCATGCTAAAAC	3468
Db	1545	pILEYASneueGlnHISGLuLYSSerAspLeuIleSerArgIleAsnGLuSerGluLY	1555

```
QY 3469 GGAAGTAGCACACTGAACATCAACACAGGTTG-----AAGGAAATAAATACCTTTGA 3522
DB 1565 SApIleGlulLeuLysSerLysLeuArgIleGlulAlaLysSerGlySerGlulLeuGl 1585
QY 3523 GGCATTAAG-----ATTTCACAGAAAG-----AATGCTGAACCTTCA 3561
DB 1585 uThrValLysGlnGlulLeuAsnAsnAlaGlnGlulLysIleArgIleAsnAlaGlulAs 1605
QY 3562 AATGACCTAAACTGAACACAGAAACAGTAACAAAAGGCGATCTCAGTATAGAGCA 3621
DB 1605 nThrValLeuLysSerLysLeuGlnLysIleArgGlnLeuLysAspLysGlnAlaGl 1625
QY 3622 GCTTAAAGTTCTGACGCGACAGACACGATGCTGACTTCTTAATTAAGGAA 3673
DB 1625 uIleLysSerAsnGlnGlulLysGlnLeuLysGlnLeuThrSerArgLeuLysGlu 1642

RESULT 6
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN [2]
RN [3]
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7899940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules.";
RL J. Biol. Chem. 275:30451-30457(2000).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z15005; CAA78727.1; -.
DR PIR; S28261; S28261.
DR HSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; -.
DR MIM; 117143; -.
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PRC0380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Alignment Scores:
Pred. No.: 2,46e-09 Length: 2663
Score: 330.00 Matches: 250
Percent Similarity: 37.53% Conservative: 224
Best Local Similarity: 19.79% Mismatches: 494
Query Match: 5.12% Indels: 295
DB: 1 Gaps: 50

US-09-602-362E-26 (1-3673) x CENE_HUMAN (1-2663)
QY 478 AATGCTTGATGGGAGGAGGACACCTCTGATGAAGGCTTACATCGAGAGGAA 537
DB 702 SerLeuIleAspGlyLysValProLysAspLeuLeuCysAsnLeuGluLeuGluGly 721
QY 538 GCTTTGTGCAATATTCTCATAGATGCTGGTGTGCTGCTAAATATTAGTAGTGATGG 597
DB 721 s-----1IleThrAspLeuGlnLysGluLeuAsn--LysGluValGlu 735
QY 598 CAACACGGCTCTCCATTATGCGGTTTATAGTGAAGATTATTATGTTGGCAACACGCT 657
DB 735 uAsnGluAlaLeuArg-----GluGluValLeuLeuLeuSerGluLeu 750
QY 658 GTCCTATGTTGTCAGTCATC-----GAGGTGCAAAACAGGCTAGCCTCAC 702
DB 750 sSerLeuProSerGluValGluArgLeuArgLysGluLeuGlnAspLysSerGluGlu 770
QY 703 ACCCTTTTACTGGCC-----ATACAGAAAGAGAGCAA 735
DB 770 uHisIleIleThrSerGluLysAspLysLeuPheSerGluValValHisLysGluSerAr 790
QY 736 G-----CAAACTGTGGAATTTTACTAAACAAAAATGC 768
DB 790 gValGlnGlyLeuLeuGluGluLeuGlyLysThrLysAspLeuAlaThrThrGln 810
QY 769 AAATGCAAAACGATTTAATGAGTCTAAATGACACAGCCCTCATGCTGCCATATGTGAAG 828
DB 810 rAsnTyLysSerThrAspGlnGluPheGlnAsnPheLysThrLeuHisMet-----As 828
QY 829 CTCATCAGATAGTCGGCATGCTTCTTCAGCAAAAGTTGACGCTCTTCTGCTGAGACAT 888
DB 828 pPheGluGlnLysTyLysMetValLeuGluGluAsnGluArgMetAsnGlnGluLeu 848
```

889 ACAT---GGAACTACGACGAGCTTATGCTGCTGCTGCTGGA-----CT 930
848 lAsnLeuSerLysGluAlaGlnLysPheAspSerLeuGlyAlaLeuLysThrGluLe 868
931 TAATTACATTCATCAACAACTTTTGGAACTATATACGAAATACCTAA----- 979
868 uSerTyLysThrGlnGluLeuGlnLysThrArgGluValGlnGluArgLeuAsnG 888
980 -----AATCCTCAAAATCAACATCCAGAGGAACATCTACAGGAACACCTGA 1026
888 uMetGluGlnLeuLysGluGlnLeuGluAsnArgAspSerProLeuGlnThrValGluAr 908
1027 TGAGGCTGCACCTTGCGCGAAGAAACACCTGACAGCGCTGAAAGCTTG-----CTGGA 1080
908 gGluLysThrLeuLeuThrGlnLysLeuGlnGlnThrLeuGluGluValLysThrLeuTh 928
1081 AAAAACACCTGACGAGCTGCACGCTTGTTGGAGGGAACGCTGCGCAAAATCAATGCT 1140
928 rGlnGluLysAspLeuLysGlnLeuGlnGlnThrLeuGlnGlnLeuGlnGlnLe 948
1141 GGGGAAGCAACATCTCGAAGTTTGACAGTCA-----ACAGAAGAACACCTAGGAA 1194
948 uLysSerAspIleHisAspThrValAsnMetAsnIleAspThrGlnGlnGlnLeuArgAs 968
1195 AATTTTGAGGCTACAAAAGAAACATCTGAGAAATTT-----TCATGGCCAGC 1242
968 nAlaLeuSerLeuLysGlnHisGlnGluThrIleAsnThrLeuLysSerLysIleSe 988
1243 AAAAGAAGATCTAGGAAGATCATCTGGGAGGAAAGAAACATCTGTAAGACTGAATG 1302
988 rGluGluValSerArgAsnLeuHisMetGluGluAsnThrGlyGluThrLysAspGluPh 1008
1303 C-----GTGGCAGGAGTAACACCTTAATAAATCTGAAGTTTGGAAAAAGGAACATC 1353
1008 eGlnGlnLysMetValGlyLe-----AspLysLysGlnAspLeuGluAlaLysAsnTh 1026
1354 TAATATGATTCATGCTCTCAAAAAGAAACATCTACAAAAGCAAGTACAAATGTGATGT 1413
1026 rGlnThrLeuThrAlaAspValLysAsp-----AsnGluIleI 1039
1414 GAGTCTGTAGAGCTATATCTAGTCTTTTGGCACACGG----- 1453
1039 eGluGlnGlnArgLysIlePheSerLeuIleGlnGluLysAsnGluLeuGlnGlnMetLe 1059
1453 ----- 1453
1059 uGluSerValIleAlaGluLysGluGlnLeuLysThrAspLeuLysGluAlaAsnIleGluMe 1079
1454 -ACTATTGAAATTCAGTGTACAAA----- 1480
1079 tThrIleGluAsnGlnGluGluLeuArgLeuLeuGlyAspGluLeuLysLysGlnGln 1099
1481 ----GTTGAGGAGACTTTAATCTTGCTACCAAGATTATCTTAAGAGTCTGCACAGAA 1536
1099 uIleValAlaGlnGluLysAsnHisAla-----IleLysLysGluGlyGluLeuSe 1116
1537 TTATACGTTTACCTGATGCTACATATCAAAAAGATATCAAAACATAAATCAAAAT 1596
1116 rArgThrCysAspArgLeuAlaGluValGluGluLysLeuLysGluLysSerGlnGlnLe 1136
1597 AGAAGATCAGATGTTCCCATCAGAACTCAAAACAGAGGAGAGTCAAGAAATATCTTGGGA 1656
1136 uGlnGluLysGlnGlnGlnLeuLeuAsnValGlnGlnGluMetSerGluMetGlnLysLy 1156
1657 TTCTGGGAGTCTC-----TTTGAGATCTCGAAAGACTCAAGTGTGTATACCTGATC 1710
1156 sIleAsnGluIleGluAsnLeuLysAsnGluLeuLysAsnLysGluLeuThrLeuGluHi 1176
1711 TATGATCAGAAAGTAAATGGAG-----ATAAATAGAGAAGTGAAGAGCTTCC 1758
1176 sMetGluThrGluArgLeuGluLeuAlaGlnLysLeuAsnGluAsnTyrcGluGluVally 1196
1759 TGAGAAGCCATCTGCTTCAAGCCTGCGGTGAAATGCAAAAGACTGTTCCAAAT----- 1813

1196 sSerIleThrLysGluArgLysValLeuLysGlnLysSerPheGluThrGluAr 1216
1814 -----AAAGCCTTTGAATTGAATGATGAATGAACAAAC 1842
1216 gAspHisLeuArgGlyTyrlleArgGluIleGluAlaThrGlyLeuGlnThrLysGluGl 1236
1843 ATTGAGACAGCTCAGATGTTTCCCATCAGAAATCCAAACAAAGACCATCAAGAAATTC 1902
1236 uLeuLysIleAlaHisIleHisLeuLysGluHisGlnGlnThrIleAspGluLeuArgAr 1256
1903 TTGGATTTCTGAGAGTCCCTGTGAGCGTTTTCACAGAAGATGTGTATTATCCCAAGC 1962
1256 gSerValSerGluLysThrAlaGlnIleIleAsnThrGlnAspLeuGluLysSerHisTh 1276
1963 TACACATCAAAAGAAATTCGATACCTTAAGTGGAAAAATTAGAAGAGTCTCTGTTAAAGA 2022
1276 rLysLeuGlnGlnGluIleProValLeuHisGluGlnGlnGlu----- 1290
2023 TCGTCTTCTGAGGCTACCTGTGGAAGGAAAGTTTCTCTCCAAATAAAGCCTTAGAATTAAAG 2082
1291 -----LeuLeuProAsnVal---LysLysValSerGluThrGlnGlnThr-----Me 1305
2083 AAAGCAGACAGAAACATTCANAGCAGAGTCTCTGATATAAGATGCTCTTCTGAAGCCTAC 2142
1305 tAsnGluLeuGluLeuLeuThrGluGlnSerThrThrLysAspSer-----Th 1321
2143 CTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAAG-----GACAG 2193
1321 rThrLeuAlaArgIleGluMetGluArgLeuArgLeuAsnGluLysPheGlnGluSerGl 1341
2194 AGAAACATCAAGCAGAGTCTCTGATGATGATGCTCTCTGAGCCTACCTCTGGAAG 2253
1341 nGluIleLysSerLeuThrLysGluArgAspAsnLeu-----LysThrIleLysGl 1359
2254 GAAAGTTTCTCTTCCAAATAAAGCTTTAGATTCAAGCAGACAGAAACATTCAAAGCAGC 2313
1359 uAlaLeuGluValLysHisAspGlnLeuLysGluHisIleArgGluThrLeu-----Al 1377
2314 TCAGATGTTTCCCATCAGAAATCCAAACAA-----AAGATGATGAAGA 2355
1377 aLysIleGlnGluSerGlnSerLysGlnGlnSerLeuAsnMetLysGluLysAspAs 1397
2356 AAATTCCTGGATTTGAGAGTCTCTGAGACTCTCTTACAGAAATGATGCTGTTTACC 2415
1397 nGluThrThrLysIleValSerGluMetGluGlnPheLysProLysAspSerAlaLeuLe 1417
2416 CAAGGCTACACATCAAAAAGAAATTCGATACCTTAAGTGGAAAAATTAGAAGTCTCTGA 2475
1417 uArgile-----GluIleGluMetLeuGlyLeuSerLysArgLeuGlnGlnSerHisAs 1435
2476 TAAAGATGCTCTCTGAGAGCCTACCTGTGGAATGAAATTTCTCTTCCAAATAAGCCTT 2535
1435 pGlu-----NectLysSerValAlaLysGluLys----- 1444
2536 AGAATTTGAAGGACAGAGAAACATTCAAAGCAGACAGATGTGAGTCTCTAGAGTCCACAT 2595
1445 -----AspAspLeuGlnArgLeuGlnGluValLe 1454
2596 CAGTCTTTTGGCAACCGACTCTGAAAATTCAGCTCTACAAAGTGTGAGGAGACTT 2655
1454 u-----GlnSerGluSerAspGlnLeuLysGluAsnIle 1465
2656 TAATCTTACTCAAGGAGGAGGCAACAAAGACAGTAACTCGACACAGGACCGTGTAT 2715
1465 e-----LysGluIleValAlaLysHisLeuGluThrGluGlnGluLysVa 1481
2716 TGGCATTATTGAACGAGTCCACAGATCAAAATAAGATGCCACATCAGATTAGG 2775
1481 lAlaHisCysCysLeuLysGluGlnGluThrIleAsnGluLeuArgValAsnLeuSe 1501
2776 AAGAAAGAGATACAAATCAACTTCAGATTCTGAGATTATCTCTGTGAGTGAT---AC 2832

Db	1501	rgLulysGluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsnAspLysLe	1521
Qy	2833	ACAGAAAT-----TATGAGTGTACTGAGCTACATATCAAAAGAAAT	2877
Db	1521	uglnAsnLysIleGlnGluIleTyrglu---LysgluGluGlnLeuAsnIleLysGlnI	1540
Qy	2878	AAAGACAAATATGCAAAATAGAGAGTCTCTGAAAAGCCTCTCACTTTGAGCCTGC	2937
Db	1540	eSerGluValGlnGluAsnValAsnGlnLysGlnPheLysGluHisArgLysAlaLy	1560
Qy	2938	CACTGAATGCAAACTGTGTTCCAAATAAAGCGTTAGAAAGGAAGATAAA-----	2989
Db	1560	sAspSerAlaLeuGlnSerIleGluSerLysMetLeuGluLeuThrAsnArgLeuGlnG	1580
Qy	2990	-----CAAACATTGAGAGCAGATTCAACTACCTATCAAAAATCTTCGA	3033
Db	1580	uSerGlnGluGluIleGlnIleMetIleLysGlnLysGluGluMetLysArgValGlnG	1600
Qy	3034	TGCACCTTCCTCTTGTGGAAGAGGAGGAACTTAAAAAGATAACTGTGAACAAATPAC	3093
Db	1600	uAlaLeu-----GlnIleGluArgAspGlnLeuLysGlnAsnThrLysGluIleVa	1617
Qy	3094	ACCAAAATGGAACAAATGCAAAAATAAG-----TTTGTGTACTA-----	3133
Db	1617	lAlaLysMetLysGluSerGlnGlnLysGluTyrglnPheLysMetThrAlaValAs	1637
Qy	3134	-----CAAAGGAACCTCTCAGAACCGAAAGAAATAAAATCACAGTTAGAACCAAAA	3186
Db	1637	ngLuthrGlnGlnLysMetCysGluIleGluHisLeuLysGluGlnPheGluThrGlnLy	1657
Qy	3187	AGCTAAATGGGAACAGAGCTCTGCAGTGTGAGATTGCTTAAATCAAGAAGAAGAA	3246
Db	1657	sLeuAsnLeuGlu-----AsnIleGluThrGluAs	1667
Qy	3247	GAGAGAAATGTCGATATATTAAAGAAAAAATTAGACCGAAGAGCAACTTAGCAAAA	3306
Db	1667	nIleArgLeuThrGlnIleGluHisGlnLeuGlnLysGluMetArgSerValThrLysGl	1687
Qy	3307	GTTAGAAGTGAACACCAACTTCAACAGACTCTCAGAATPACAGATATAGAAATGAAA--	3364
Db	1687	uArgAspLeuArgSerValGluGluThrLysValGluArgAspGlnLysGlnG	1707
Qy	3365	-----AGTGTACAGTAATTTGAAAT-----CAGTTTCTCA	3396
Db	1707	uAsnLeuArgGluThrIleThrArgAspLeuGlnLysGlnGluLeuLysIleValHi	1727
Qy	3397	CACTCATGAAAGTGAATAAT-----	3415
Db	1727	sMetHisLeuLysGluHisGlnGluThrIleAspLysLeuArgGlyIleValSerGluLy	1747
Qy	3416	-----GATCTCTTTGATAAATTTGATCTGATTTGAAAAA	3447
Db	1747	sThrAsnGluIleSerAsnMetGlnLysAspLeuGluHisSerAsnAspAlaLeuLysAl	1767
Qy	3448	GGAATTCATGCTCAAACTGGAAGTACCACTGAAACATCAACACAGGTGAAGGA	3507
Db	1767	aGlnAspLeuLysIleGlnGluLeu---ArgIleAlaHisMetHisLeuLysGluG	1786
Qy	3508	AAATAAATACTTTGAGGACATT-----	3529
Db	1786	nGlnGluThrIleAspLysLeuArgGlyIleValSerGluLysThrAspLysLeuSerAs	1806
Qy	3530	-----NAGATTTACAGAAAGATGCTGAATTTCAA-----	3562
Db	1806	nMetGlnLysAspLeuAsnSerAsnAlaLysLeuGlnGluLysIleGlnGluLeuLy	1826
Qy	3563	-----ATGACCTTAAACTGAAACAGAAACAGAAACAGTAAACAAAGGCGC	3603
Db	1826	sAlaAsnGluHisGlnLeuIleThrLeuLysLysAspValAsnGluThrGlnLysLysVa	1846
Qy	3604	ATCTAGTATAGAGAGCTTAAAGTTCTGACCGCAGAGACACATGCTGACTCTTAA	3663
Db	1846	lSerGlu---MetGlnGlnLeuLysGlnIleLysAspGlnSerLeuThrLeuSerLy	1865

Qy	3664	ATTGAAG	3670
Db	1865	sLeuGlu	1867
RESULT 7			
MYHB	MOUSE	STANDARD;	PRT; 1972 AA.
ID	MYHB	MOUSE	
AC	008638;	008639;	062462;
DT	16-OCT-2001	(Rel. 40,	Created)
DT	16-OCT-2001	(Rel. 40,	Last sequence update)
DT	15-MAR-2004	(Rel. 43,	Last annotation update)
DB	Myosin	heavy chain,	smooth muscle isoform (SMMHC).
GN	MYH11.		
OS	Mus musculus	(Mouse).	
OC	Eukaryota;	Metazoa;	Chordata;
OC	Mammalia;	Eutheria;	Rodentia;
OC			Sciurognathi;
OC			Muridae;
OC			Murinae;
OC			Mus.
OK	NCBI	taxid=10090;	
RN	[1]	SEQUENCE FROM N.A.	
RP	STRAIN=BALB/c;	TISSUE=Uterus;	
RC	MEDLINE=97242182;	PubMed=9125171;	
EX	Hasegawa K.,	Arakawa E.,	Oda S.,
RA	Matsuda Y.,		
RT	"Molecular cloning and expression of murine smooth muscle myosin heavy		
RT	chains.";		
RL	Biochem. Biophys. Res. Commun.	232:313-316	(1997).
RN	[2]	SEQUENCE OF 1-368 FROM N.A.	
RP	TISSUE=Uterus;		
RC	MEDLINE=95008063;	PubMed=7923625;	
EX	Miano J.M.,	Cserjesi P.,	Ligon K.L.,
RA	Periasamy M.,	Olson E.N.;	
RT	"Smooth muscle myosin heavy chain exclusively marks the smooth muscle		
RT	lineage during mouse embryogenesis.";		
RL	Circ. Res.	75:803-812	(1994).
RN	[3]	SEQUENCE OF 1-126 FROM N.A.	
RP	TISSUE=Uterus;		
RC	MEDLINE=96172919;	PubMed=8593698;	
EX	Suzuki T.,	Kim H.S.,	Kurabayashi M.,
RA	Hamada H.,	Fujii H.,	Aikawa M.,
RT	Watanabe M.,	Watanabe N.,	Sakomura Y.,
RT	Nagai R.;		
RT	"Preferential differentiation of P19 mouse embryonal carcinoma cells		
RT	into smooth muscle cells. Use of retinoic acid and antisense against		
RT	the central nervous system-specific POU transcription factor Brn-2.";		
RL	Circ. Res.	78:395-404	(1996).
CC	-!- FUNCTION: Muscle contraction.		
CC	-!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2		
CC	heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)		
CC	and 2 regulatory light chain subunits (MLC-2).		
CC	-!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.		
CC	-!- ALTERNATIVE PRODUCTS:		
CC	Event-Alternative splicing; Named isoforms=2;		
CC	Name=1;		
CC	Isoid=008638-1; Sequence=Displayed;		
CC	Name=2;		
CC	Isoid=008638-2; Sequence=VSP_003346;		
CC	-!- DOMAIN: The rodlike tail sequence is highly repetitive, showing		
CC	cycles of a 28-residue repeat pattern composed of 4 heptapeptides,		
CC	characteristic for alpha-helical coiled coils.		
CC	-!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light		
CC	meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be		
CC	split further into 2 globular subfragments (S1) and 1 rod-shaped		
CC	subfragment (S2).		
CC	-!- SIMILARITY: Contains 1 myosin-like globular head domain.		
CC	-!- SIMILARITY: Contains 1 IQ domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		


```
QY 2027 CTTCTGAAGCCTACCTGTCGAAGGAAAGTTTCTCTCCAAATAAAGACCTTAGAATTA--- 2083
Db 1300 -----GluGlyLeuAlaIleLysLeuAla 1307
QY 2084 AAGCAGCAGAGAAACATTCAAAGCAGAGTCTCTCTGATAAAGATGGTCTCTGGAAGCTACC 2143
Db 1308 LysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGlnLeuLeuGlnGluGlu 1327
QY 2144 TGTGGAAGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAGACAGCAGAGAAACATC 2203
Db 1328 ThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSer 1347
QY 2204 AAAGCAGAGTCTCTCATATATGATGTCTCTGAAGCCTACCTGTGGAGGAAAGTTTCT 2263
Db 1348 LeuGlnAspGlnLeuAspGluMetGluAlaLysGlnAsnLeuGluArgHisValSer 1367
QY 2264 CTTCCAAATTAAGCTTTAGATTGAAGAGCAGAGAAACATCAAAGCAGCTCAGATGTC 2323
Db 1368 ThrLeuAsnIleGlnLeu----- 1373
QY 2324 CCATCAGAAATCCAAACAAAGGATGATGAAGAAATCTTGGGATTTTGAAGTTTCCTT 2383
Db 1374 ---SerAspSerLysLys-----LeuGlnAspPheAlaSerThrIle 1387
QY 2384 GAGACTCTCTTACAGAAATGATGTGTCTTACCAGGCTACACATCAAAAGAAATTCGAT 2443
Db 1388 GluValMetGluGluGly-----LysLysArgLeuGlnLysGluMetGlu 1402
QY 2444 ACCTTAAGTGGAAATTAAGAG-----TCTCTCATATAA 2479
Db 1403 GlyLeuSerGlnGlnThrGluGluLysAlaAlaTyrAspLysLeuGluLysThrLys 1422
QY 2480 GATGGTCTCTGAAGCCTACCTGTGGAATGAAATTTCTCTCCAAATAAAGCC----- 2533
Db 1423 AsnArgLeuGlnGlnGluLeuAspLeuValValAspLeuAspAsnGlnArgGlnLeu 1442
QY 2534 -----TTGAAATTGAAGACAGA----- 2551
Db 1443 ValSerAsnLeuGluLysLysGlnLysLysPheAspGlnLeuLeuAlaGluLysAsn 1462
QY 2552 -----GAAACATTCAAAGCAGAGGATGTGAGTTCTGTAGAGTCC 2590
Db 1463 IleSerSerLysTyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGlu 1482
QY 2591 ACATTCAAGTCTTTTGGCAACCGACTACTGAAATTTCAAGTCTACAAAGTTGAGGAA 2650
Db 1483 ThrLysAlaLeuSerLeuAlaArgAlaLeuGlu-----GluAlaLeuGluAlaLysGlu 1500
QY 2651 GACTTTAATCTTACTACCAAGAG-----GGACCAACAAG 2686
Db 1501 GluLeuGluArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLys 1520
QY 2687 ACAGTAACCTGGACAAACAGAAACGTGATATTGGCAATTATTGAACGAGCTCCAAAGATCAA 2746
Db 1521 AspAspValGlyLysAsnValHisGluLeuGluLysSerLysArgAlaLeuGluThrGln 1540
QY 2747 ACAAAATAGATGCCCATCAGAAATTTAGGAAGAAAGAGATACAAATCAATCTCAGAT 2806
Db 1541 MetGluGluMetLysThrGlnLeuGluGluGluSerGluAspValGlnAlaThrGluAsp 1560
QY 2807 TCTGAGATT---ATCTCTGTGAGTGATACACAGAAATTTAGTGTGTACCTGAGCTACA 2863
Db 1561 AlaLysLeuArgLeuGluVal-----AsnMetGlnAlaLeu---LysGlyGln 1575
QY 2864 TATCAAAAAGAAATAAGACAAACAAATGGCAAAATAGAGAGTCTCTCT-----GAA 2914
Db 1576 PheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArgArgGlnLeuGln 1595
QY 2915 AAGCCTTCTCAGCTTTCAGCTGCACTGAAATGCAAAACTCTGTTCCTCAAAATTAAGGCTTA 2974
Db 1596 ArgGlnLeuHis---GluTyrGluThrGluLeuGluAspGluArgLysGlnArgAlaLeu 1614
```

```
QY 2975 GAATGGAAGATAAACAACATTTGAGAGCAGATTCACTACCTCATCAAAAATCTTTGGAT 3034
Db 1615 AlaAlaAlaAlaLysLysLysLeuGluGlyAspLeuLysAspLeuGluLeuGlnAlaAsp 1634
QY 3035 GCACTTCTCTTGTGAAAGAGGAGGAACTTTAAAAAAGATAACTCTGTGAACAAATTTACA 3094
Db 1635 -----SerAlaIleLysGlyArgGlu-----GluAlaIleLysGlnLeuArg 1648
QY 3095 GCAAAATATGGNACAAATGAAATAAGTTTGTGTACTACAAAGAGAACTGTGAGAAGCG 3154
Db 1649 LysLeuGlnAlaGlnMetLysAsp-----PheGlnArgGluLeuAspAla 1664
QY 3155 AAAGAAATAAAATCACAGTTA-----GAGAACCAAAAA-----GCTAAA 3193
Db 1665 ArgAlaSerArgAspGluIlePheAlaThrSerLysGluAsnGluLysLysAlaLysSer 1684
QY 3194 TGGGAACAAGAGCTCTCGAGTGTGAGATTCGCTTTAATCAAGNAGAGAGAGAGAGA 3253
Db 1685 LeuGluAlaAspLeuMetGlnLeuGlnGluAspLeuAlaAlaGluArgAlaArgLys 1704
QY 3254 AATCTCGATATATTAAAGAAAAAATTAGACCCGAA----- 3289
Db 1705 GlnAlaAspLeuGluLysGluGluLeuAlaGluLeuAlaSerSerLeuSerGlyArg 1724
QY 3290 -----GAGCAACTTAGGAAAAAGTTAGAACTGAAA---CACCAACTTTGACAGACT 3337
Db 1725 AsnThrLeuGlnAspGluLysArgArgLeuGluAlaArgIleAlaGlnLeuGluGluGlu 1744
QY 3338 CTCGAATACAAGATATAGAAATTGAAAAGTGTA-----ACAAAGT 3376
Db 1745 LeuGluGluGluGlnGlyAsnMetGluAlaMetSerAspArgValArgLysAlaThrLeu 1764
QY 3377 AATTGTAATCAGGTTTCTCACACTCATCAAGAGTGAATAATGATCTCTTTTCATGAAAT--- 3433
Db 1765 GlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArgSerThrAlaGlnLysAsnGlu 1784
QY 3433 ----- 3433
Db 1785 SerAlaArgGlnGlnLeuGluArgGlnAsnLysGluLeuArgSerLysLeuGlnGluVal 1804
QY 3434 -----TGCATGTTGAAAAAGAAATTCGCCATGTGTAATAACTGGAAGTAGCC 3478
Db 1805 GluGlyAlaValLysAlaLysLeuLysSerThrValAlaAlaLeuGluAlaLysIleAla 1824
QY 3479 ACATGTAACATCAACACCGAGTGAAGAAAAATAATATCTTTGAGGACATTAGATTTTA 3538
Db 1825 GlnLeuGluGlnValGluGlnGluGlnAlaArgGluLysGlnAlaAlaThrLysSerLeu 1844
QY 3539 CARGAAAAGATCTGCAACTTCAAAATGCACCTCAAACTGAAACAGAAAAACAGTAACAAAA 3598
Db 1845 LysGlnLysAspLysLysLeuLysGluValLeu---LeuGlnValGluAspGluArgLys 1863
QY 3599 AGGCATCTCAGTATAGAGAGCAGCTTAAAGTTCTGACGCGCAGAGAACACGATGCTGACT 3658
Db 1864 MetAlaGluGlnTyrLysGluGlnAlaGluLysGlyAsnThrLysValLysGlnLeuLys 1883
QY 3659 TCTAAATTAAGGAA 3673
Db 1884 ArgGlnLeuGluGlu 1888
RESULT 8
MYH9 HUMAN
ID MYH9 HUMAN STANDARD; PRT; 1960 AA.
AC P35579; O60805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057185; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clump M., Smink L.J., Alincough R., Almeida J.P., Babbage A.K.,
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloydd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wansley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Filahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [2]
 RP SEQUENCE OF 1-1337 FROM N.A.
 RX MEDLINE=92003925; PubMed=1912569;
 RA Tothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
 RA Arnaout M.A., Clayton L.K., Tennen D.G.;
 RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
 RT cDNA clones, characterization of the protein, chromosomal
 RT localization, and upregulation during myeloid differentiation.";
 RL Blood 78:1826-1833(1991).
 RN [3]
 RP SEQUENCE OF 1-715 FROM N.A.
 RX MEDLINE=91316803; PubMed=1860190;
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Gula D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located
 RT on different chromosomes.";
 RL Circ. Res. 69:530-539(1991).
 RN [4]
 RP SEQUENCE OF 714-1960 FROM N.A.
 RX MEDLINE=90138958; PubMed=1967836;
 RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
 RN [5]
 RP VARIANT DFNA17 HIS-705.
 RX MEDLINE=20489856; PubMed=11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Whare A.N.;
 RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
 RT nonmuscle myosin MYH9.";
 RL Am. J. Hum. Genet. 67:1121-1128(2000).
 RN [6]
 RP VARIANTS MHA/FTNS/SBS LYS-93; CYS-702; CYS-1165; HIS-1424 AND
 RP LYS-1841.
 RX MEDLINE=20428192; PubMed=10973259;
 RA Seri X., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Chiggeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
 RA Iglason A., Zelan L.L., Savoia A., Balduini C.L., Norris P.,
 RA Magrini U., Belletti S., Heath K.E., Babcock M., Gluckman M.J.,
 RA Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [7]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE=20428193; PubMed=10973260;
 RA Kelley M.J., Jawien W., Ortel T.L., Korczak J.F.;
 RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
 CC cell shape, and specialized functions such as secretion and
 CC capping.
 CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
 CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
 CC regulatory light chain subunits (MLC-2).
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -!- DISEASE: Defects in MYH9 are the cause of May-Hegglin anomaly
 CC (MHA) [MIM:155100]. MHA is an autosomal dominant
 CC macrothrombocytopenia characterized by thrombocytopenia, giant
 CC platelets and leukocyte inclusions.
 CC -!- DISEASE: Defects in MYH9 are the cause of Fechtner syndrome (FTNS)
 CC [MIM:153640]. FTNS is an autosomal dominant macrothrombocytopenia
 CC characterized by thrombocytopenia, giant platelets and leukocyte
 CC inclusions. With additional alport-like clinical features of
 CC sensorineural deafness, cataracts and nephritis.
 CC -!- DISEASE: Defects in MYH9 are the cause of Sebastian syndrome (SBS)
 CC [MIM:605249]. SBS is an autosomal dominant macrothrombocytopenia
 CC characterized by thrombocytopenia, giant platelets and leukocyte
 CC inclusions.
 CC -!- DISEASE: Defects in MYH9 are the cause of autosomal dominant
 CC nonsyndromic sensorineural deafness 17 (DFNA17) [MIM:603622].
 CC DFNA17 is characterized by progressive hearing impairment and
 CC cochleosaccular degeneration.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; Z82215; -; NOT ANNOTATED_CDS.
 DR EMBL; M81105; AAA59888.1; -;
 DR EMBL; M69180; AAA61765.1; -;
 DR EMBL; M31013; AAA36349.1; -;
 DR PIR; A61231; A61231.
 DR HSSP; P10587; 1BR2.
 DR Genew; HGNC:7579; MYH9.
 DR MIM; 155100; -;
 DR MIM; 155100; -;
 DR MIM; 153640; -;
 DR MIM; 605249; -;
 DR MIM; 605249; -;

DR MM; 603622; -
DR GO; GO:0005860; C:non-muscle myosin; TAS.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; Myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PD00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;
KW Multigene family; Disease mutation; Deafness.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 837 1926 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT VARIANT 93 93 N -> K (IN MHA).
FT /FTID=VAR 010791.
FT VARIANT 702 702 R -> C (IN FTNS).
FT /FTID=VAR 010792.
FT VARIANT 705 705 R -> H (IN DFNAL7).
FT /FTID=VAR 010793.
FT T -> I (IN MHA).
FT /FTID=VAR 010794.
FT R -> C (IN SRS).
FT /FTID=VAR 010795.
FT D -> H (IN FTNS).
FT /FTID=VAR 010796.
FT E -> K (IN MHA).
FT /FTID=VAR 010797.
FT EAI -> RGH (IN REF. 3).
FT CONFLICT 53 55 T -> S (IN REF. 3).
FT CONFLICT 660 660 T -> M (IN REF. 4).
FT CONFLICT 869 869 C -> Y (IN REF. 4).
FT CONFLICT 931 931 KG -> GR (IN REF. 4).
FT CONFLICT 1240 1241 E -> BE (IN REF. 2).
FT CONFLICT 1350 1350

Alignment Scores:
Pred. No.: 8 44e-09 Length: 1960
Score: 319.00 Matches: 219
Percent Similarity: 38.23% Conservative: 187
Best Local Similarity: 20.62% Mismatches: 434
Query Match: 4.95% Indels: 222
DB: 1 Gaps: 40

US-09-602-362E-26 (1-3673) x MYH9_HUMAN (1-1960)

QY 854 CTTGAGAAAGTTTGAACAGTCAACA-----GAAGAAACACCTAGAAATTT 1198
Db 998 GluAspArgIleAlaGluPheThrThrAsnLeuThrGluGluGluLysSerLysSer 1017
QY 1199 TTGAGGCTACAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCTAGG 1258
Db 1018 LeuAlaLysLeuLysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArg 1037
QY 1259 AAGATCATCATGGGAGGAGAAACATCTGTAAGACTGAATCGCTGGCAGGAGTAAACA 1318
Db 1038 Arg-----GluGluLysGlnArgGlnGluLeuLys-----Thr 1049
QY 1319 CCTAATAAACTGAGTTTGGGAAAAGGACATCTAATATGATTGCA----- 1366
Db 1050 ArgArgLysLeuGluGlyAspSerThrAspLeuSerAspGlnIleAlaGluLeuGlnAla 1069
QY 1367 -----TGTCCTACAAAGAAACATCTACAAAGCAAGT 1399
Db 1070 GlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaLeu 1089
QY 1400 ACAATGTGATGTGAGTCTCTGTAGAGCCTATATTCAGTCTCTTTTGGCAGGACTATT 1459
Db 1090 AlaArgValGluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeu 1109
QY 1460 GAAATTCACAGTGTACAAAGTTGAGGAAGACTTTAATCTTGCTACCAAGATTCTCT 1519
Db 1110 Glu---SerGlnIleSerGluLeuGlnLysLeuGlu-----Ser 1122
QY 1520 AAGAGTCTGCACAGAATTATACGTGTTTACCTGATGCTACATATCAAAAGATATCAAA 1579
Db 1123 GluArgAlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuGluGluGluGlu 1142
QY 1580 ACATAAATCACAAATAGAGATCATGTTCCCATCA-----GAA 1621
Db 1143 AlalaLysThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGluLeuArg 1162
QY 1622 TCCAAGCAGAGGAAGATGAAGATATTTCTGGGATCTCTGGAGTCTCTTTGAGAGTCT 1681
Db 1163 SerLysArgGluGlnGluValAsnIle-----LeuLysLysThrLeuGluGluGlu 1179
QY 1682 GCAAGACTCAAGTGTGTATCTACCTGAGTCTATGTATCAGAAAGTAATGGAGATAAATAGA 1741
Db 1180 AlalysThrHisGluAlaGlnIleGlnGluMetArgGlnLysHisSerGlnAlaValGlu 1199
QY 1742 GAAGTAGAAGAGTCTCTGAGAGCCATCTGCCTTCAAGCTGCGGTNGAAATGCAAAAG 1801
Db 1200 GluLeuAlaGlnLeuGluGlnThrLysArgValLysAlaAsnLeuGluLysAlaLys 1219
QY 1802 ---ACTGTTCCAAATAAAGCCTTTGAATGAAGAAATCAACAAACATTGAGAGAGCTCAG 1858
Db 1220 GlnThrLeuGluAsnGluArgGlyGluLeuAlaAsnGluValLysValLeuLeuGlnGly 1239
QY 1859 ATGTTCCCATCAGATCCAAACAAAGGACGATGAGAAATTTCTGGGATTTCTGAGAGT 1918
Db 1240 LysGlyAspSerGluLysLysArgLysValGluAlaGlnLeuGlnGluLeuGlnVal 1259
QY 1919 CCCTGT---GAGACGGTTTCCACAGAGGATGTGTATTTACCCAAAGCTACACAT---CAA 1972
Db 1260 LysPheAsnGluGlyGluArgValArgThrGluLeuAlaAspLysValThrLysLeuGln 1279
QY 1973 AAGAATTCGATACCTTAAGTGGAAATTTAGACAGTCTCTGTTAAGATGGTCTTCTG 2032
Db 1280 ValGluLeuAspAsnValThrGlyLeuLeuSerGlnSerAspSerLysSerLysLeu 1299
QY 2033 AAGCCTACTCTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAGGACAGA 2092
Db 1300 -----ThrLysAspPhe 1303
QY 2093 GAACATTCAAGCAGAGTCTCTCGATAAAGATGGTCTTCTGAAGCCTACTCTGGAGAG 2152
Db 1304 SerAlaLeuGluSerGlnLeuGlnAspThrGlnGluLeuGlnGluGluAsnArgGln 1323

```
QY 2153 AAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAGAC-----AGAGAA 2197
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1324 LysLeuSerThrLysLeuLysGlnValGluAspGluLysAsnSerPheArgGlu 1343
QY 2198 ACACCTCAAGCAGAGCTCTCGTATATGATGGTCTTCTGAG-----CCTACCTGTGA 2251
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1344 GlnLeuGluGluGluGluGluAlaLysHisAsnLeuGluLysGlnLeuAlaThrLeuHis 1363
QY 2252 AGGAAGTTCTCTTCCAAATAAAGCTTTAGAA-----TTGAAGGACAGA 2296
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1364 AlaGlnValAlaAspMetLysLysLysMetGluAspSerValGlyCysLeuGluThrAla 1383
QY 2297 GAACATTCAGAGCAGCTCAGATGTTCCCATCAGAA-----TCCAAACAAGAGATGAT 2350
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1384 GluGluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGlnArgHisGluGlu 1403
QY 2351 GAAGAAAATCTTGGGATTTGAGAGTTCTTGGAGACTCTCTTACAGAAAT-----GAT 2404
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1404 LysValAlaAlaTyrispLysLeuGluLysThrLysThrArgLeuGlnGlnLeuAsp 1423
QY 2405 GTGTGTTTACCAAGGCTACACAT-----CAA 2431
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1424 AspLeuValAspLeuAspHisGlnArgGlnSerAlaCysAsnLeuGluLysLysGln 1443
QY 2432 AAAGAAATTCGAT-----ACCTTAAGTGGAAAAATTAGAAGATCT 2470
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1444 LysLysPheAspGlnLeuAlaGluGluLysThrIleSerAlaLysTyrispAlaGluGlu 1463
QY 2471 CCTGATAAAGATGCTCTTCTGAAGCCTACCTGTGGAATGAAATTTCTCTTCCAAATAAA 2530
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1464 ArgAspArg-----Ala 1467
QY 2531 GCCTTAGAANTGAAGNACAGACAAACATTCAGACAGAGATGTAGTCTGTAGATGCC 2590
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1468 GluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGluGlu 1487
QY 2591 ACATTGAGTCTTTTGGCAACCGACTACTGAAATTCAGACTCTACAAAAGTTGAGGAA 2650
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1488 AlaMetGluGlnLysAlaGluLeuGluArgLeuAsnLysGlnPheArgThrGluMetGlu 1507
QY 2651 GACTTAAATCTTACTACAGAGGAGGAGCACAACACAGACTGACTGGACACAGGAACGT 2710
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1508 AspLeu---MetSerLysAspValGlyLysSerVal-----His 1521
QY 2711 GATATTGGCATTTATGAACAGCTCCACAGATCAACAAATAAGATGCCACATCAGAA 2770
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1522 GluLeuGluLysSerLysArgAlaLeuGluGlnGlnValGluMetLysThrGlnLeu 1541
QY 2771 TTAGGAAGAAAGAGATACAAATCAACTTCAGATCTGAGATTT---ATCTCTGTGAGT 2827
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1542 GluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluVal--- 1560
QY 2828 GATACACAGATTTATGAGTGTTTTACCTGAGGCTACATATCAAAAAGATAAGACAA 2887
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1561 -----AsnLeuGlnAlaMet---LysAlaGlnPheGluArgAspLeuGln----- 1574
QY 2888 AATGGCAAAATAGAGAGTCTCTGAAAAGCCCTTCTCAGTTT-----GAG 2932
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1575 ---GlyArgAspGluGlnSerGluGluLysLysLysGlnValArgGlnValArgGlu 1593
QY 2933 CTGCGCACTCAATGCAAAACTCTGTTCCAAATAAAGGCTTAGATGGAAGATAACAA 2992
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1594 MetGluAlaGluLeuGluAspGluArgLysGlnArgSerMetAlaValAlaAlaArgLys 1613
QY 2993 ACATTGAGCAGATTCACACTACCTCCATCAAAAATCTTGGATGCA----- 3037
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1614 LysLeuGluMetAspLeuLysAspLeuGluAlaHisIleAspSerAlaAsnLysAsnArg 1633
QY 3038 -----CTTCCTCTCTTGTGAAGAGGA 3058
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1634 AspGluAlaIleLysGlnLeuArgLysLeuGlnAlaGlnMetLysAspCysMetArgGlu 1653
QY 3059 AGCGAACTTAAAAAGATACTCTGTGAACAAATTACAGCA----- 3097
```

```
Db 1654 LeuAspSerThrArgAlaSerArgGluGluLeuAlaGlnAlaLysGluAsnGluLys 1673
QY 3098 AAAATGGACAAATACAAAATAAAGTTTGTCTACTACAAAAGGAACCTGTGAGAGCGAAA 3157
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1674 LysLeuLysSerMetGluAlaGluMetIleGlnLeuGlnGluLeuAlaAlaGlu 1693
QY 3158 GAAATAAATCAGCTAGTAGACACCAAAAGCTTAATGGGAACAAGAGCTCTGCAAGTGTG 3217
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1694 ArgAlaLysArgGlnAlaGlnGlnGluArgAspGluLeuAlaAspGluIleAlaAsnSer 1713
QY 3218 -----AGATTGCCTTTAAATCAAGAAGAAGAGAGAGAAATTCGTGATATA 3265
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1714 SerGlyLysGlyAlaLeuAlaLeu-----GluGluLysArgArg----- 1726
QY 3266 TTAAGAAAGAAAATTT---AGACCCGAGAGCAACTTAGGAAAAGTTAGAGTGAACAC 3322
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1727 LeuGluAlaArgIleAlaGlnLeuGluGluGluLeuGluGlnGlnGlyAsnThrGlu 1746
QY 3323 CACTTGAACAGACTCTCAGATACAGATATAGAATTGAAAGTCTAAACAGTAATTTG 3382
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1747 LeuIleAsnAspArgLeuLysLysAlaAsnLeuGlnIleAspGlnIleAsnThrAspLeu 1766
QY 3383 AAT---CAGGTTTCTCAGCTCATGAAAGTGAATAATGATCTCTTTCATGAAAATTGCATG 3439
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1767 AsnLeuGluArgSerHisAlaGlnLysAsnGluAsn-----AlaArgGlnGln 1782
QY 3440 TTGAAAAGGAAATGCGATGCTTAAACTGGAAGTAGCCACTGAAACATCAACACCAG 3499
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1783 LeuGluArgGlnAsnLysGluLeuLysValLys-----LeuGlnGluMetGluGly 1799
QY 3500 GTGAAGAAATAATAATCTTGAGCAGATTAAAGATTTTACAAGAAAAGATGCTGAACCT 3559
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1800 ThrValLysSerLysTyrispLysAlaSerIleThrAlaLeuGluAlaLysIleAlaGlnLeu 1819
QY 3560 CAATGACCCCTA----- 3571
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1820 GluGluGlnLeuAspAsnGluThrLysGluArgGlnAlaAlaCysLysGlnValArgArg 1839
QY 3572 -----AACTGAAACAGAAACAGTA-----ACAAAAGGGCATCT 3607
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1840 ThrGluLysLysLysLysLysValLeuLeuGlnValAspAspGluArgArgAsnAlaGlu 1859
QY 3608 CAGTATAGAGCAGCTTAAAGTTCTGACGGCAGAGAACACGATGCTGACTCTTAATTTG 3667
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1860 GlnTyrispAspGlnAlaAspLysAlaSerThrArgLeuLysGlnLeuLysArgGlnLeu 1879
QY 3668 AAGGAA 3673
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1880 GluGlu 1881

RESULT 9
MYHB HUMAN
ID MYHB HUMAN STANDARD; PRT; 1972 AA.
AC P35749; O00396; O94944; P78422;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH1 OR KIAA0866.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
```

RL Genomics 60:295-308 (1999).
 RN [2] SEQUENCE FROM N.A., AND REVISIONS TO 1263-1266.
 RP TISSUE=Brain;
 RC Nagase T., Kikuno R., Yamakawa H., Ohara O.;
 RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [3] SEQUENCE OF 1-1266 FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RX Matsuo K. R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:355-364 (1998).
 RN [4] SEQUENCE OF 885-1972 FROM N.A.
 RP MEDLINE=93263189; PubMed=7684189;
 RA Matsuo K. R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
 RX Yanagisawa M., Masaki T., Takao A.;
 RA "Human smooth muscle myosin heavy chain gene mapped to chromosomal
 RT region 16q12.";
 RL Am. J. Med. Genet. 46:61-67 (1993).
 RN [5] SEQUENCE OF 1093-1972 FROM N.A.
 RP TISSUE=Hippocampus;
 RC Okajima K.;
 RA Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Muscle contraction.
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- TISSUE SPECIFICITY: Smooth muscle; expressed in the umbilical
 CC artery, bladder, esophagus and trachea.
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -!- DISEASE: A chromosomal rearrangement, known as pericentric
 CC inversion Inv(16)(p13;q22), produces a fusion protein that
 CC consists of the 165 N-terminal residues of CBF-beta (PEPB2) with
 CC the tail region of MYH11. This rearrangement is associated with
 CC acute myeloid leukemia of M4EO subtype.
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/MYH11ID43.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF001548; AAC31665.1; -;
 CC EMBL; U91323; AAC35212.2; -;
 CC EMBL; AB020673; BAA74889.2; ALT INIT.
 CC EMBL; D10667; -; NOT ANNOTATED_CDS.
 CC EMBL; X69292; CAA49154.1; -;
 CC HSSP; P10587; 1BR2.
 CC Genew; HGNC:7569; MYH11.
 CC MIM; 160745; -;
 CC GO; GO:0005859; C:muscle myosin; TAS.
 CC InterPro; IPR000048; IQ region.
 CC InterPro; IPR001609; myosin head.
 CC InterPro; IPR004009; Myosin_N.

DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family;
 KW Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 1 785
 FT DOMAIN 786 815
 FT DOMAIN 844 1934
 FT DOMAIN 1935 1972
 FT NP_BIND 178 185
 FT DOMAIN 661 683
 FT DOMAIN 762 776
 FT MOD_RES 129 129
 FT CONFLICT 887 889
 FT CONFLICT 1558 1558
 FT CONFLICT 1610 1611
 FT CONFLICT 1786 1786
 FT CONFLICT 1958 1958
 FT CONFLICT 1972 AA; 227338 MW; 67665BB2AECE1277 CRC64;
 SQ SEQUENCE 1972 AA; 227338 MW; 67665BB2AECE1277 CRC64;
 Alignment Scores:
 Pred. No.: 8,44e-09 Length: 1972
 Score: 319.00 Matches: 244
 Percent Similarity: 36.82% Conservative: 186
 Best Local Similarity: 20.89% Mismatches: 438
 Query Match: 4.95% Indels: 300
 DB: 1 Gaps: 49
 US-09-602-362E-26 (1-3673) x MYHB_HUMAN (1-1972)
 QY 704 CCCCTTTTACTGGCCATACAGAAAGCAAGCAAACTGTG-----GAATTTTA 754
 DB 843 ProLeuLeuGlnValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGln 862
 QY 755 CTACAAAAAATGCAATGCAAGCGCTTAATGAGCTAAATGCAAGCGCTTATGCT 814
 DB 863 LysThrLysGluArgGlnGlnLysAlaGluAsnGluLeuLysGluLeuGlnLysHis 882
 QY 815 GCCATATGTGAAGGCTCATCAGAGATAGTCGGCATGCTTCTTCAGCAAAATGTTAGCTC 874
 DB 883 SerGlnLeuThrGluGluLysAsnLeuLeuGlnGluGlnLeuGlnAlaGluThrGluLeu 902
 QY 875 TTTGCTGAAGACATACATGGAATAACTGCGAGAA-----CGTTATGCTGCTGCT 922
 DB 903 TyrAlaGlu-----AlaGluGluMetArgValArgLeuAlaLys 916
 QY 923 CGTGGAGTTAATACATTCATCACTTTCGACATATACGAAATACCTAAAAAT 982
 DB 917 Lys-----GlnGluLeuGluLeuLeuHisGluMetGluAlaArg 930
 QY 983 CCTCAAAATACCAATCCAGGAAGCAACATCTACAGGAAACCTCATGAGGCTGCACCTTG 1042
 DB 931 LeuGluGluGluGluAspArgGlyGlnGlnLeuGlnAla-----GluArgLysLysMet 948
 QY 1043 GCGGAAAGAACACCTGACCGGCTGAAGCTTGTGGAAGAAACACCTGACGAGCTGCA 1102
 DB 949 AlaGlnGlnMetLeuAspLeuGluGlnLeuGluGlu-----GluGluAlaAla 965
 QY 1103 CGCTTGGTGGAGGGAACGCTCTGCCAAATTCATATGCTGGGAAAGCAACATCTGGAAG 1162
 DB 966 Arg-----GlnLysLeuGln---LeuGluLysValThrAla----- 976
 QY 1163 TTTGAAACAGTCAACAGAAAGAACACCTGAGAAATTTTGGAGCTTACAAAAGAACATCT 1222

Db 977 ---GluAlaLysIleLysLysLeuGluAspGluLeu----- 988
QY 1223 GAGAAATTTTCATGGCCAGCAAAAGAAAGATCTAGGAAGATCACATGGGAGGAAAGAA 1282
Db 989 -----ValMetAspAspGlnAsnAsnLysLeuSer-----LysGlu 1000
QY 1283 ACATCTGTAAGACTCAATCGCGGAGGAGTAACACACTTAATAAACTGAAGTTTGGAA 1342
Db 1001 ArgLysLeuGluGluGluArgIleSerAspLeuThrThrAsnLeuAlaGluGluGlu 1020
QY 1343 AAGGAAACATCTAATATGATTCATCTCCTCAAAAGAAACATCTACAAAACAGATACA 1402
Db 1021 Lys---AlaLysAsnLeuThrLysLeuLysAsnLysHisGluSerMetIleSerGluLeu 1039
QY 1403 AATGTG----- 1408
Db 1040 GluValArgLeuLysLysGluGluLysSerArgGlnGluLeuGluLysLysArgLys 1059
QY 1409 -----GATGTGATTCGTGA---GAGCCTATATTCAGTCTTTTGGCCACACGGACT 1456
Db 1060 LeuGluGlyAspAlaSerAspPheHisGluGlnIleAlaAspLeuGlnAlaGlnIleAla 1079
QY 1457 ATTGAAATTCACAGGTACAAAGTTGAGGAGACTTAACTTGTCTACCAAGATTATC 1516
Db 1080 GluLeuLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaAlaLeuAlaArgLeu 1099
QY 1517 TCTAAGAGTGTCTGCACAGAAATTATACGTCTTTACCTGATGCTACATATCAAAAAGATATC 1576
Db 1100 AspAspGluIleAlaGlnLysAsnAlaLeu-----LysLysIle 1113
QY 1577 AAAACAATAATCACAATAATAGAGATCAGATGTTCCCATCAGATCCAAA----- 1627
Db 1114 ArgLeuLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAspSerGluAlaAla 1133
QY 1628 CGAGAGCAGATGAGAATATTCTTGGGATCTTGGGAGTCTCTTGGAGAGTTCTGCAAG 1687
Db 1134 ArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGlu---AlaLeuLys 1152
QY 1688 ACTCAAGTG-----TGATACCTAGTCTATGTATCAGAAAGTAATGAGATAAATAGA 1741
Db 1153 ThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnGluLeuArgAlaLysArg 1172
QY 1742 GAA----- 1744
Db 1173 GluGlnGluValThrValLeuLysLysAlaLeuAspGluGluThrArgSerHisGluAla 1192
QY 1745 -----GTAAAGAGCTTCTCTGAGAGCCA 1768
Db 1193 GlnValGlnGluMetArgGlnLysHisAlaGlnAlaValGluGluLeuThrGlnGlnLeu 1212
QY 1769 TCTGCCTTCAAGCCTCCGTGNGAAATGCAAGACTGTTCCAAATAAAGCCTTTGAAATG 1828
Db 1213 GluGlnPheLysArgAla-----LysAlaAsnLeuAsp 1223
QY 1829 AAGAATCAACAACATTTG----- 1846
Db 1224 LysAsnLysGlnThrLeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeu 1243
QY 1847 ---AGAGCAGCTCAGATGTTCCATCAGATCCAAACAAAGCAGCATGAGAAATTTCT 1903
Db 1244 GlyGlnAlaLysGln-----GluValGluHisLysLysLysLeuGluAlaGlnVal 1261
QY 1904 TGGGATCTGAGAGTCCCTGTGAGCGGTTTCCACAGAGGATGTGATTTACCCAAAGCT 1963
Db 1262 GlnGluLeuGlnSerLysCysSerAspGlyGluArgAlaArgAlaGluLeuAsnAspLys 1281
QY 1964 ACACAT-----CAAAAGAAATTCGATCTTAAGTGGAAATTTAGAGAGTCTCCTGTT 2017
Db 1282 ValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAla----- 1299
QY 2018 AAAGATGGTCTTCTGAAGCCTACCTGTGGAAGAAAGATTCTCTTCCAAATAAAGCCTTA 2077
Db 1300 -----GluGlyLysAlaIle 1304

QY 2078 GAATTA---AAGGACACAGAAACATTCAAAGCAGAGTCTCTCTGATAAAGATGGTCTTCTG 2134
Db 1305 LysLeuAlaLysAspValAlaSerLeuSerSerGlnLeuGlnAspThrGlnGluLeuLeu 1324
QY 2135 AAGCTTACCTGTGGAGGAAAGTTTCTTCTCCAAATAAAGCCCTTAGAATTAAGGACAGA 2194
Db 1325 GlnGluGluThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluGluGlu 1344
QY 2195 GAAACACTCAAGCAGAGTCTCTGATAATGATGGTCTTCTGAAGCCTACCTCTGGAAGG 2254
Db 1345 ArgAsnSerLeuGlnAspGlnLeuAspGluGluMetGluAlaLysGlnAsnLeuGluArg 1364
QY 2255 AAGTTTCTCTTCCAAATAAAGCTTTAGAAATTTGAAGCAGACAGAAACATTCAAAAGCAGCT 2314
Db 1365 HisIleSerThrLeuAsnIleGlnLeu----- 1373
QY 2315 CAGATGTTCCATCAGATCCAAACAAAGGATGATCAAGAAATTTCTTGGGATTTTGAG 2374
Db 1374 -----SerAspSerLysLysLys-----LeuGlnAspPheAla 1384
QY 2375 AGTTTCTCTGAGACTCTCTTACAGAAATGATGTGTGTACCCAGGCTACACATCAAAA 2434
Db 1385 SerThrValGluAlaLeuGluGly-----LysLysArgPheGlnLys 1399
QY 2435 GAATTCGATACCTTAACTGGAATAATTAGAAG-----TCT 2470
Db 1400 GluIleGlnAsnLeuThrGlnGlnTyrgluGluLysAlaAlaAlaTyraAspLysLeuGlu 1419
QY 2471 CCTGATAAAGATGTCTTCTGAAGCCTACCTGTGGATGAAATTTCTTCTCCAAATAAA 2530
Db 1420 LysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuValValAspLeuAspAsnGln 1439
QY 2531 GCC-----TTAGAATTGAAGGACAGA----- 2551
Db 1440 ArgGlnLeuValSerAsnLeuGluLysLysGlnArgLysPheAspGlnLeuLeuAlaGlu 1459
QY 2552 -----GAAACATTCAAAGCAGAGGATGAGTTCT 2581
Db 1460 GluLysAsnIleSerSerLysTyraAspGluArgAspArgAlaGluAlaAlaArg 1479
QY 2582 GTAGATCCACATTCAGTCTTTTGGCAACCGACTACTGAAATTCACAGTCTCAAAA 2641
Db 1480 GluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGlu-----GluAlaLeuGlu 1497
QY 2642 GTTCGAGAGACTTTAATCTTACTTACCAGGAG-----GGA 2677
Db 1498 AlaLysGluGluLeuGluArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuVal 1517
QY 2678 GCAACAAAGACAGTAACCTGGCAACAGGAACTGATATTGGCATTATTGCAACGAGCTCCA 2737
Db 1518 SerSerLysAspValGlyLysAsnValHisGluLeuGluLysSerLysArgAlaLeu 1537
QY 2738 CAAGATCAACAATAAGATGCCCATCATCAGATTTAGGAAGAAAGAGATACAATCA 2797
Db 1538 GluThrGlnMetGluGluMetLysThrGlnLeuGluGluLeuAspGluGluGlnAla 1557
QY 2798 ACTTCAGATCTCAGATT---ATCTCTGTGAGTGATACACAGAAATTATGAGTGTTCACCT 2854
Db 1558 ThrGluAspAlaLysLeuArgLeuGluVal-----AsnMetGlnAlaLeu--- 1572
QY 2855 GAGCTCATATCAAAAGAAATAAAGACAAACAAATGGCAAAATAGAGAGTCTCCT--- 2911
Db 1573 LysGlyGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArgArg 1592
QY 2912 -----CAAAAGCCTTCTCCTTTGAGCCTGCCACTGAAATGCAAACTCTGTTCAAAAT 2965
Db 1593 GlnLeuGlnArgGlnLeuHis---GluTyrgluThrGluLeuGluAspGluArgLysGln 1611
QY 2966 AAGGCTTAGAATGGAAGATAAACAACATTTGAGAGCAGATTCAACTACCTTATCAAAA 3025
Db 1612 ArgAlaLeuAlaAlaAlaLysLysLysLeuGluGluLysAspLeuLysLeuLeu 1631


```

QY 3026 ATCTTGATGCATCTCTCTTGTGAAGAGGAGGAACTTAAAGAGATAACTGTGAA 3085
D 3027 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1632 GlnAlaAsp-----SerAlaIleIysGlyArgGlu-----GluAlaIleIys 1645
QY 3086 CAAATTACAGCAAAATGAAACAAATGAAATAGTTTGTGTACTACAAAGAACTG 3145
D 3087 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1646 GlnLeuArgLysLeuGlnAlaGlnMetIysAsp-----PheGlnArgGluLeu 1661
QY 3146 TCAGAGCGGAGAAATAAATACAGTTA-----GAGAACCCMAAA--- 3187
D 3147 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1662 GluAspAlaArgAlaSerArgAspGluIlePheAlaThrAlaLysGluAsnGluLys 1681
QY 3188 ---GCTAAATGGGAACAAGAGCTGCTGAGTGTGAGATTGCCCTTAAATCAAGAAGAG 3244
D 3189 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1682 AlaLysSerLeuGluAlaAspLeuMetGlnLeuGlnGluAspLeuAlaAlaGluArg 1701
QY 3245 AAGAGAGAATGTGATATATATAAAGAAAAAATAGACCCGAA----- 3289
D 3246 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1702 AlaArgLysGlnAlaAspLeuGluLysGluGluLeuAlaGluGluLeuAlaSerSerLeu 1721
QY 3290 -----GAGCAACTTAGCAAAAGTTAGAGAGTGA---CACCAACTT 3328
D 3291 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1722 SerGlyArgAsnAlaLeuGlnAspGluLysArgGluGluAlaArgIleAlaGlnLeu 1741
QY 3329 GAACAGACTCTCAGATACAGATATAGAAATGTAAGTGTAAAGTAATTTGAATCAG 3388
D 3330 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1742 GluGluGluLeuGluGluGlnGlyAsnMetGluAlaMetSerAspArgValArgLys 1761
QY 3389 GTTCTCACACTCATGAAAGT---GAAATGATCTC----- 3421
D 3390 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1762 AlaThrGlnGlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArgSerThrAlaGln 1781
QY 3422 -----NATTCATGTTGAAAAGCAAAATGCCATGCTTAAACTG 3469
D 3423 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1802 HisGluMetGluGlyAlaValLysSerLysPheLysSerThrIleAlaLeuGluAla 1821
QY 3470 GAAGTAGCCACACTGAAACATCAACACACAGGTGGAAGGAAAAATAATCTTTGAGGACATT 3529
D 3471 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1822 LysIleAlaGlnLeuGluGlnValGlnGlnGluAlaArgLysGlnAlaAlaThr 1841
QY 3530 AGAATTTACAGAAAGAAATCTGAACTTCAAACTCAACCTTAAACTGAAACAGAAACA 3589
D 3531 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1842 LysSerLeuLysGlnLysAspLysLysLysLysLysLysLysLysLysLysLysLys 1860
QY 3590 GTAACAAAAGGCACTCTCAGTATAGAGAGCAGTAAAGTTCTGAGCGCAGAGAACACG 3649
D 3591 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1861 GluArgLysMetAlaGluGlnThrLysGluGlnAlaGluLysGlyAsnAlaArgValLys 1880
QY 3650 ATGCTGACTTCTAAATTGAAGGAA 3673
D 3651 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1881 GlnLeuLysArgGlnLeuGluGlu 1888

RESULT 10
MYS2_DICDI
ID MYS2_DICDI STANDARD; PRT; 2116 AA.
AC P08739;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
ON NCBI_TaxID=44689;
RX SEQUENCE FROM N.A.
RP MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Levinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from

```

```

RT Dictyostelium discoideum."
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Luck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
  Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
  phosphorylatable heavy chain fragment of Dictyostelium myosin II."
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
  Dictyostelium myosin heavy chain."
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
  Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
  discoideum complexed with MgADP.Befx and MgADP.ALf4-."
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
  truncated head of Dictyostelium discoideum myosin to 2.7-A
  resolution."
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
  Dictyostelium discoideum myosin motor domain to 1.9-A resolution."
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammAS, and MgAMPNP complexes
  of the Dictyostelium discoideum myosin motor domain."
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
  of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
  Dictyostelium discoideum myosin motor domain."
RL J. Mol. Biol. 274:394-407(1997).
CC -!- FUNCTION: Myosin is a protein that binds to actin and has ATPase
  activity that is activated by actin.
CC -!- SUBUNIT: Myosin II heavy chain is two-headed. It self-assembles
  into filaments. Hexamer of 2 heavy chain subunits (MHC), 2 alkali
  light chain subunits (MLC) and 2 regulatory light chain subunits
  (MLC-2).
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
  CORTEX.
CC -!- DOMAIN: Each myosin heavy chain can be split into 1 light
  meromyosin (LM) and 1 heavy meromyosin (HMM). It can be further
  split into 2 globular subfragments (S1) and 1 rod-shaped
  subfragment (S2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
  cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
  characteristic for alpha-helical coiled coils.
CC -!- PTM: Phosphorylation inhibits thick filament formation and reduces
  the actin-activated ATPase activity.

```


QY	1196	ATTTGAGGCCTACAAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAATCTCT	1255
DB	1303	::: :::: ----LysLysGlnLysGluSerAsnGluLys-----	1311
QY	1256	AGGAAGATCACATGGGAGGAAAAGAAACATCTGTAAGACTGAATCGCTGGCAGGAGTA	1315
DB	1312	ArgLysValAspLeuGluLysGluValSerGluLeuLysAspGlnIleGluGluGluVal	1331
QY	1316	ACACCTAATAAAA-----ACTGAAGTTTTCGAAAAAGGAAACATCTAATATGATTGCATGT	1369
DB	1332	AlaserLysLysAlaValThrGluAlaLysAsnLysLysGluSerGluLeuAspGluIle	1351
QY	1370	CCTACAAAAGAAACATCTACAAAAGCAAGTACAAATGTGGATGTGAGTTCTAGAGCCT	1429
DB	1352	LysArgGlnTyrAlaAspValValSerSerArgAsp-----LysSerValGluGln	1368
QY	1430	ATATTACAGTCTTTTGGCACACGACTATTGCAAAATTCACAGTGTACAAAGTTTGAGGAA	1489
DB	1369	LeuLysThrLeuGlnAlaLysAsn-----GluGluLeuArgAsnThrAlaGluAla	1386
QY	1490	GACTTAAATCTTGCTACCAAGATTATCTTAAGAGTCTGCACAGATTATACGTGTTTA	1549
DB	1387	GluGlyGlnLeuAspArgAlaGluArgSerLysLysLysAla-----	1400
QY	1550	CCTGATCTACATATCAAAAAGATATCAAAACATAAATCACAAAATAGAAGATCAGATG	1609
DB	1401	--GluPheAspLeuGluAlaValLysAsnLeuGluGluGlnThrAlaLysVal	1419
QY	1610	TTCCCATCAGATCCAAACGAGAGGAAGATGAGATATTTCTTGGGATTCTGGAGTCTC	1669
DB	1420	LysAlaGluLysAlaMetLysLysAlaGluThrAspTyrArgSerThrLysSerGluLeu	1439
QY	1670	TTTGAGAGTTCTGCAAAAGACTCAAGTGTGTACCTGAGTCTATGTATCAGAAAGTAATG	1729
DB	1440	-----AspAspAlaLysAsn-----ValSerSerGluGlnTyrValGlnIleLys	1454
QY	1730	GAGATAAATAGAAAGTAGAAGAG-----CTTCTCAGAAAGCCATCTGCCTTC	1777
DB	1455	ArgLeuAsnGluGluLeuSerGluLeuArgSerValLeuGluGluAlaAspGluArgCys	1474
QY	1778	AAGCCTCCCTGNGAAATGCCAAAGACATGTTCCAAATAAAGCCTTGCAATTGAAAGATGAA	1837
DB	1475	AsnSerAlaIleLysAlaLysLysThrAlaGluSerAlaLeuGluSerLeuLysAspGlu	1494
QY	1838	CAAAATAGAGCAGCTCAGATGTTCCCATCAGATCCAAACAAAGGACGATGAA---	1894
DB	1495	-----IleAspAlaAlaAsnAsnAlaLysAlaGluArgLysSerLysGluLeu	1512
QY	1895	-----GAAATTCITGGGATTCGAGAGT-----CCCTGT	1924
DB	1513	GluValArgValAlaGluLeuGluGluSerLeuGluAspLysSerGlyThrValAsnVal	1532
QY	1925	GAGACGGTTTCACAGAGGATGTGTATTTACCCAAAGCTACACATCAAAGAAATTCGAT	1984
DB	1533	GluPheIleArgLysLysAspAla-----GluIleAsp	1543
QY	1985	ACCTTAAGTGAAATTAGAA-----GAGTCTCCTGTTAAAGATGCTCTCTGAAG	2035
DB	1544	AspLeuArgAlaArgLeuAspArgGluThrGluSerArgIleLysSerAsp-----	1560
QY	2036	CCTACCTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTTAAAGGACAGAA	2095
DB	1561	-----GluAspLysLys	1564
QY	2096	ACATCAAAAGCAGAGTCTCCTGATAAAGATGGTCTCTGAAGCCTACCTGTGGAAGGAAA	2155
DB	1565	AsnThrArgLysGlnPheAlaAspLeuAla-----Lys	1576
QY	2156	GTTTCTCTTCCAAAATAAAGCCCTTAGAATTTAAAGGACAGAGAAACACATCAAGCAGAGTCT	2215
DB	1577	ValGluGluAlaGlnArgGluValValThrIleAspArgLeuLysLysLysLeuGluSer	1596

2216 CTGATTAATGAT-----GGTCTTCTGAAGCCCTACCTGTGGAAG---AAAGTTTCT 2266
Db |||||
1597 AspIleAspLeuSerThrGlnLeuAspThrGluThrLysSerArgIleIysIleGlu 1616
Qy |||||
2264 CTTCCAAATAAAGCTTTAGAA-----TTCAAGACACAGAGAAAACATCAAGACAGAGCTCAG 2317
Db |||||
1617 LysSerLysLysLeuGluGlnThrLeuAlaGluArg-----ArgAlaAlaGlu 1633
Qy |||||
2318 ATGTTCCTCATCAGATCCAAACAAAAGGATGATGAAGAAAATTTCTGGGATTTTGAGAGT 2377
Qy |||||
1634 GluGlySerLysAlaAlaAspGluGluIleArgLysGlnValTrpGlnGluValAsp 1653
Db |||||
2378 TTCTTCAGACTCTCTACAGATGATGTGTGTTTACCCAAAGCTACACATCAAAA--- 2434
Qy |||||
1654 GluLeuArgAlaGlnLeuAspSerGluArgAlaAlaLeuAsnAlaSerGluLysLysIle 1673
Db |||||
2435 -----GAATTTCGATACCTTAAGTCGAAAATTAGAA---GAGTCTCCTGAT 2476
Qy |||||
1674 LysSerLeuValAlaGluValAspGluValLysGluGlnLeuGluAspGluIleLeuAla 1693
Db |||||
2477 AAGATCGTCTTCGAGCCCTACTCTGCGAATGAAATTTCTTCCCAATATAAGCCTTA 2536
Qy |||||
1694 LysAspLysLeuValLysAlaLysArgAlaLeuGluValGlnLeuGluValArgAsp 1713
Db |||||
2537 GAATTGAAGACACAGAGAAACATTCAAAGCAGAG----- 2569
Qy |||||
1714 GlnLeuGluGlnGluGluAspSerArgSerGluLeuGluAspSerLysArgGluThr 1733
Db |||||
2570 ---GATGTGAGTCTGTAGAGTCCCATTCAGTCTTTTTCGCAACCGACTACTGAAAAT 2626
Qy |||||
1734 ThrGluValGluAspIleLysLysLysArgAsp-----AlaGluVal 1747
Db |||||
2627 TCAGAGTCTCAAAAGTTGAGGAAGACTTTAATCTTACTCAAGGAGGAGCAACAAG 2686
Qy |||||
1748 GluGlnAsnThrLysLeuAspGlu-----AlaLysLys 1758
Db |||||
2687 ACAGTAATCGACACAGAAACGTGATATTGCGCATTTATGAACGAGCTCCACAGATCAA 2746
Qy |||||
1759 LysLeuThrAsp-----AspValAspThrLeuLysLysGlnLeuGluAspGlu 1774
Db |||||
2747 ACAATAGATGCCCATCATCAGATATTAGGAGAAA-----GAAGAT 2788
Qy |||||
1775 LysLysLysLeuAsnGluSerGluArgAlaLysLysArgLeuGluSerGluAsnGluAsp 1794
Db |||||
2789 ACAAAATCAACTTCAGATTCCTGAGTATCTCTGTGATGATACACAGAAATTATGAGTGT 2848
Qy |||||
1795 PheLeuAlaLysLeuAspAlaGluValLysAsnArgSerArgAlaGluLys----- 1811
Db |||||
2849 TTACTGAGCTACATATCAAAAGAATAAAGACACAAATGCCAAATAGAAGAGTCT 2908
Qy |||||
1812 ---AspArgLysLysTrpGluLysAspLeuLysAspThrLysTrpLysLeuAsnAsp--- 1829
Db |||||
2909 CCTGAAAAGCCCTCTCACTTTGAGCTGCCACTGAAATGCAAAACTCTGTCCAAATAAA 2968
Qy |||||
1830 -----GluAlaAlaThrLysThrGlnThrGluIleGlyAlaAla 1842
Db |||||
2969 GCGTTAGATGGAAGATAAACAACATTCAGACAGATTCACACTACCTATCAAAATC 3028
Qy |||||
1843 LysLeuGlu----- 1845
Db |||||
3029 TTGATGCACTTCCTTCTTGTGAAAGAGGAGAACTTAAAAAAGATAACTGTGAACAA 3088
Qy |||||
1846 -----AspGlnIleAspGlu 1850
Db |||||
3089 ATTACAGCAAAATGGAACAAATGAAAATAAGTTTTGTGTACTACAAAAGGAAGTCTCA 3148
Qy |||||
1851 LeuArgSerLysLeuGluGlnGluAlaLysAlaThrGlnAlaAspLys----- 1867
Db |||||
3149 GAACGGAAGAAATAAATCACACTTAGAGAACCAAAAGCTAATCGGCAACAGAGCTC 3208
Qy |||||
1868 SerLysLysThrLeuGluGluGluIleAspAsnLeuArgAlaGlnIleGluAspGlu--- 1886
Db |||||
3209 TCGAGTGTGAGATTGCCTTTAAATCAAGAAGAGAGAGAGA-----AATGTC 3259
Qy |||||

984 GluAspAspIleLeuValMetAspAspGlnAsnAsnLys----- 996
1019 ACACCTGATGAGCTGCACCTTGGCGGAAGAACACCTGACACGGCTGAAGCTTGCTG 1078
997 LeuSerLysGluArgLysLeuLeuGluGluArgGileSerAspLeuThrAsnLeuAla 1016
1079 GAAAAACACCTGACGAGCTGCACGCTTGGTGGAGGGAACGCTGCCAAAATTCATGCT 1138
1017 GluGlu---GluGluLysAlaLysAsnLeuThrLys----- 1027
1139 CTGGGGAAGCAACATCTCGAAGATTGTAAGCTCAACAGAGAACAACCTAGGAATTT 1198
1028 -----LeuLysAsnLysHisGluSerMetIleSerGluLeuValArgLeu 1043
1199 TTGAGCCTCAAAAACATCTGAGAAATTTTCATGCCAGCAAAAGAAAGATCTAGG 1258
1044 LysLysGluLysSerArgGlnGlu-----LeuGluLysLysArg 1058
1259 AAGATCACATGGAGGAAAAAGAACATCTGTAAAG---ACTGAATCGGTGGCAGAGTA 1315
1059 LysMetAspGlyGluAlaSerAspLeuHisGluGlnIleAlaAspLeuGlnAlaGlnIle 1078
1316 ACACCTAATAAACTGAAGTTTGGNAAGGAACATCTAATATGATTCATGCTCTACA 1375
1079 AlaGluLeuLysMetGlnLeuAlaLysLysGluGluLeuGlnAlaLeuAlaArg 1098
1376 AAGAAGAACATCTACAAAACAAAGTACAAATGTGGATGTGAGTTCTGTAGAGCCTATATC 1435
1099 LeuGluAspGluThrSerGlnLysAsnAsn----- 1108
1436 AGTCCTTTGGCACCGGACTATTGAAATTCACAGTGTACAAAGTTGAGGAAGACTTT 1495
1109 AlaLeuLysLysIleArgGluLeuGlu---GlyHisIleSerAspLeuGlnGluAspLeu 1127
1496 AATCTTGCTACCAAGATTATCTTAAGAGTGTGCACAGAAATATACGTGTTTACCTGAT 1555
1128 Asp-----SerGluArgAlaAlaArgAsnLysAlaGluLysGlnLys 1141
1556 GTTACATATCAAAAGATATCAAAACAATAAATCAAAATAGAAGATCAGATGTTCOCA 1615
1142 ArgAspLeuGlyGluGluLeuGluAlaLysThrGluLeuGluAspThrLeuAspThr 1161
1616 TCAGATCAACACGAGAG-----GAAGATGAAGATATCTTGGGATTCGGG 1663
1162 ThrAlaThrGlnGlnLeuAlaLysArgGluGlnGluValThrValLeuLysLys 1181
1664 AGTCCTTTGAGAGTTCTGCAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAA 1723
1182 AlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGlu---MetArgGlnLys 1200
1724 GTAATGGAGATAAATAGAGAAGTAGAGAGCTTCTGAGAACCCATCTGCCTCAAGCCT 1783
1201 HisThrGlnVal-----ValGluGluLeuThrGluGlnLeuGluGlnPheLysArg 1217
1784 GCGGTN-----CAATGCAAAAGACTGTTCACAAATAAAGCCCTTGAATGAAG 1831
1218 AlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLysGluAsnAlaAspLeuAla 1237
1832 AATGAACAACATAGAGACGCTCAGATGTTCCCATCAGATCCAAACAAAGACCGAT 1891
1238 GlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLysLeu 1257
1892 GAAGAAAATCTTGGGATTCAGAGTCCCTGAGACGGTTTCACAGAGGATGTGTAT 1951
1258 GluValGlnLeuGlnLeuGlnSerLysCysSerAspGlyGluArgAlaArgAlaGlu 1277
1952 TTACCCAAAGCTACACAT-----CAAAAGAAATTCGATACCTTAAGTGGAAAATAGAA 2005
1278 LeuAsnAspLysValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuSer 1297
2006 GAGTCTCTGTTAAAGATGCTTCTGAGAGCCTACTCTGGAGGAAAGATTCTCTTCCA 2065
|||||

1298 GluAla----- 1300
2066 AATAAGCCTTAGAATTA---AAGACAGAGAAACATTCAAAGCAGAGTCTCCTGATAAA 2122
1301 GlyLysAlaIleLysLeuAlaLysGluValAlaSerLeuGlySerGlnLeuGlnAspThr 1320
2123 GATGCTCTTCTGAAAGCCTACCTGTGGAGGAAGAAATTTCTCTCCAAATAAAGCCCTAGAA 2182
1321 GlnGluLeuLeuGlnGluThrArgGlnLysLeuAsnValSerThrLysLeuArgGln 1340
2183 TTAAGGACAGAGAAACACTCAAGCAGAGTCTCCTGATATGATGTCTTCTGAAGCCT 2242
1341 LeuGluAspGluArgAsnSerLeuGlnGlnLeuAspGluGluMetGluAlaLysGln 1360
2243 ACCTGTGGAAGAAATTTCTCTCCAAATAAAGCTTTAGAAATTAAGGACAGAGAAACA 2302
1361 AsnLeuGluArgHisIleSerThrLeuAsnIleGlnLeu----- 1373
2303 TTCAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACAAAGAGATGATGAAGAAAATCT 2362
1374 -----SerAspSerLysLysLys-----Leu 1380
2363 TGGGATTTGAGAGTTCCTTGAGACTCTTACAGAAATGATGTGTGTTACCCCAAGCT 2422
1381 GlnAspPheAlaSerThrValGluSerLeuGluGly-----LysLys 1395
2423 ACACATCAAAAGAAATTCGATACCTTAAGTGGAAAATTAGAAGAG---TCTCCTGATAAA 2479
1396 ArgPheGlnLysGluIleGluSerLeuThrGlnGlnLysAlaAlaLys 1415
2480 GATGCTCTTCTGAGCCTACCTGTGGAATGAAAATTTCTCTCCAAATAAGCCTTAGAA 2539
1416 AspLysLeuGluLysThrLysAsnArgLeuGlnGlnLeuAspLeuValValAsp 1435
2540 TTGAAGCAGACAGAAACATTCAAAGCAGAGATGTGAGTCTGTAGAGTCCACATTCAGT 2599
1436 LeuAspAsnArgGlnLeu-----ValSerAsnLeuGluLysLysGlnLys 1451
2600 CTTTTTGGCAACCGACTACTGAAAATTCACAG---TCTACAAAAGTTGAGGAAGAC--- 2653
1452 LysPheAspGlnLeuLeuAlaLysGluLysAsnIleSerSerLysThrAlaAspGluArg 1471
2654 -----TTTAATCTTACTACCAAGGAGGAGCAACAGACAGTAACCTGGACACACGAA 2707
1472 AspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAla 1491
2708 CGTGATATTGGCAATTATTGAACGAGCTCCACAGATCAAAACAAATAAGATGCCACATCA 2767
1492 LeuGluGluAlaLeuGluAlaLysGluGluLeuGluArgThrAsnLysMetLeuLysAla 1511
2768 GAATTAGGAAGAAAGAGATACAAAATCAACTTCAGATTCTGAGATTATCTCTGTGAGT 2827
1512 GluMet-----GluAspLeuValSerSerLysAspValGlyLysAsnValHis 1528
2828 GATACACAGAAATATGATGTTTACCTGAGCTACATATCAAAAGAAATAAAGACAA--- 2884
1529 GluLeuGluLysSerLysArgAlaLeuGlu---ThrGlnMetGluGluMetLysThrGln 1547
2884 ----- 2884
1548 LeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluVal 1567
2885 -----ACAAATGCAAAATAGAGAGTCT 2908
1568 AsnMetGlnAlaLeuLysValGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsn 1587
2909 CCTGAAAGCCTTCTCACTTT-----GAGCCTGCCACTGAAATGCCAAAC 2953
1588 GluGluLysArgArgGlnLeuGlnArgGlnLeuHisGluLysThrGluLeuGluAsp 1607
2954 TCTGTTCCAAATAAAGCCTTAGAATGGAAGAAATAAACAACATTGAGAGCAGATTCACCT 3013
1608 GluArgLysGlnArgAlaLeuAlaAlaLysLysLysLeuGluGlyAspLeuLys 1627
|||||

2261 TCTCTTCCAAATAAAGCTTTAGAAATTGATGAGGACAGAGAAACATTCAGAGCTCAGATG 2320
Db : : : : :
303 GlnSer-HisLysGluGlnCysThrLeuLeuThrSerGlu- 315
QY 2321 TTCCCATCAGATCCAAACAAAGGATGATGAGGAAATCTTGGGATTTGAG- 2374
Db : : : : :
316 -----LysGluAlaLeuGlnGluGlnLeuAspGluArgLeuGlnGluGluLysIle 333
QY 2375 -----AGT 2377
Db : : : : :
334 LysAspLeuHisMetAlaGluLysThrLysLeuLeuThrGlnLeuArgAspAlaLysAsn 353
QY 2378 TTCCTTGAGACTCTCTTACAGATGATGCTGTTTACCCAGGCTACATCAAAAGAA 2437
Db : : : : :
354 LeuIleGluGlnLeuGlnAspLysGlyMetValIleAlaGluThrLysArgGlnMet 373
QY 2438 TTCGATACCTTAAGTGGAAAATTAGAAGAGTCTCTGTGATAAAGATGCTTCTGAG- 2494
Db : : : : :
374 HisGluThrLeuGluMetLysGluGluGluIleAlaGlnLeuArgSerArgIleLysGln 393
QY 2495 ---CCTACCTGTGGAATGAAATTTCTCTCCAAATAAAGCCTTAGAATTGAGCAGACA 2551
Db : : : : :
394 MetThrThrGlnGlyGluGluLeu-----ArgGluGlnLysGluLysSerGluArg 410
QY 2552 GAAACATTCAAACGAGGATGAGTCTGTGATAGTCCACATTCAGTCTTTTGGCAA 2611
Db : : : : :
411 AlaAlaPhe-----GluGluLeuGluLysAlaLeuSerThr-----AlaGln 424
QY 2612 CCAGTACTGAAATTCACAGTCTACAAAGTTGAGGAAGCTTTAACTTACTACC- 2668
Db : : : : :
425 LysThrGluGluAlaArgLysLeuLysAlaGluMetAspGluGlnIleLysThrIle 444
QY 2669 ---AAGGAGGAGCAACAAGACAGTAACCTGGACACAGGACGCTGAT- 2713
Db : : : : :
445 GluLysThrSerGluGluGluArgIleSerLeuGlnGlnGluLeuSerArgValLysGln 464
QY 2714 -----ATTGGCATATTGACAGAGCTCCACAGATCAA-----ACAAATAGATG 2758
Db : : : : :
465 GluValValAspValMetLysLysSerSerGluGluGlnIleAlaLysLeuGlnLysLeu 484
QY 2759 CCACATCAGAAATTAGAGAAAGAGAGATACAAATCAACTTCAGATTCTCAGATTATC 2818
Db : : : : :
485 HisGluLysGluLeuAlaArgLysGlu-----GlnGluLeuThr 497
QY 2819 TCTGTGAGTGATACAGAAATTATGAGTGTATTACCTGAGGCTACATATCAAAAGAAATA 2878
Db : : : : :
498 LysLysLeuGlnThrArg-----GluArgGluPheGlnGluGlnMet 511
QY 2879 AAGACAAATAATGCAAAATAGAGAGTCTCTGTAAGAGCCTTCTCAGTTTGAGCTGCC 2938
Db : : : : :
512 LysValAla-----LeuGluLysSerGlnSerGluThrLysLysIleSerGlnGlu 528
QY 2939 ACTGAAATGCAAACTCTGTTCCAAATAAAGGCTTAGAATGGAAGAAATAAACAACATG 2998
Db : : : : :
529 LysGluGlnGlnGluSerLeuAlaLeuGluGluLeuGluGlnLysLysAlaIleLeu 548
QY 2999 AGAGCAGATTCACTACCTCA-----TCAAA 3025
Db : : : : :
549 ThrGluSerGluAsnLysLeuArgAspLeuGlnGlnGluAlaGluThrThrArgThrArg 568
QY 3026 ATCTTGATGACTCTCTCTTGTGAGAGGAGGAACTTAAAGAAATCACTGTGAA 3085
Db : : : : :
569 IleLeuGlnLeuGluSerLeuGluLysSerLeuGlnGlnLysAsnLysGlnSerLys 588
QY 3086 CAAATTCAGCAAAATGGAACAAATGAAAATAAGTTT-----TGTGTA 3130
Db : : : : :
589 AspLeuAlaValHisLeuGluAlaGluLysAsnLysHisAsnLysGluIleThrValMet 608
QY 3131 CTACAAAGGAACGTGTCAGAGCGGAAGAATAAATCAAGTTAGAGAACCAAAAGCT 3190
Db : : : : :
609 ValGluLysHisLysThrGluLeuGluSerLeuLysGln-----GlnAspAla 625

3191 AAATGGCAACAAGAGCTCTGACGTGAGATGCTCTTTAAATCAAGAAGAAGAG- 3247
Db : : : : :
626 LeuThrThrGluLysLeuGlnValLeuGlnGlnThrGlnThrGlnMetGluLysLeu 645
QY 3248 AGAAGAAATGTCGATATATTAAAGAAAATTTAGACCGAGAGCAACTTAGGAAAAAG 3307
Db : : : : :
646 ArgGluLysCysGluGlnGluLysGluThrLeuLeuLysAspLysGluIleIlePheGln 665
QY 3308 TTAGAAGTGAACACCAACTTGAACAGACTCTCAGAAATACAGATATA-----GAA 3358
Db : : : : :
666 AlaHisIleGlnGluMetAsnGluLysThrLeuGluLysLeuAspValLysGlnThrGlu 685
QY 3359 TTGAAAAGTGTAAACAGTAAATTTGAATCAGGTTTTCACACTCATGAAAAGTGAATGAT 3418
Db : : : : :
686 LeuGluSerLeuSerGluLeuSerGluValLeuLysAlaArgHisLys----- 702
QY 3419 CTCCTTTCATGAAATTCATGTTGAAAAGGAATTTGCCATGCTAAACTGGAAGTAGCC 3478
Db : : : : :
703 -----LeuGluGluGluLeuSerValLeuLysAspGlnThrAsp 715
QY 3479 ACACCTGAAACATCAACACAGGTCAAG-----GAAATAAATACCTTTGAGGAC 3526
Db : : : : :
716 LysMetLysGlnGluLeuGluAlaLysMetAspGluGlnLysAsnHisGlnGlnGln 735
QY 3527 ATT---AAGATTTTACAGAAAAGATGCTGAATCTCAATGACC----- 3568
Db : : : : :
736 ValAspSerIleLysGluHisGluValSerIleGlnArgThrGluLysAlaLeuLys 755
QY 3569 -----CTAAAACCTGAAACACAGACACAGTACAAAAGGCGCATCTCAGTAT 3613
Db : : : : :
756 AspGlnIleAsnGlnLeuGluLeuLeuLysGluArgAspLysHisLeuLysGluHis 775
QY 3614 AGACAGAGCTTAAAGTTCTGACGGCAGAG 3643
Db : : : : :
776 GlnAlaHisValGluAsnLeuGluAlaAsp 785

RESULT 14
MYH9_RAT STANDARD; PRT; 1961 AA.
AC Q62812;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
type A) (Nonmuscle myosin heavy chain-A) (NMMEC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
cell shape, and specialized functions such as secretion and
capping.
CC !- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
regulatory light chain subunits (MLC-2).
CC !- DOMAIN: The rodlike tail sequence is highly repetitive, showing
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
characteristic for alpha-helical coiled coils.
CC !- SIMILARITY: Contains 1 myosin-like globular head domain.
CC !- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```

CC EMBL; U31463; AAA74950.1; -.
DR HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00936; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
SQ SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;

Alignment Scores:
Pred. No.: 2,77e-08 Length: 1961
Score: 308.50 Matches: 227
Percent Similarity: 38.36% Conservative: 198
Best Local Similarity: 20.49% Mismatches: 454
Query Match: 4.78% Indels: 229
DB: 1 Gaps: 43

```

US-09-602-362E-26 (1-3673) x MYH9_RAT (1-1961)

```

QY 704 CCCCTTTTACTGCCATACAGAAAAGAACGACCAAACTGTGGAATTTTACTACAAAA 763
DB 836 ProLeuLeuAsnSerIleArgHisGluAspGlu-----LeuLeuAlaLys 850
QY 764 AATGMAATGCAACGATTTAATGAGTCTAATGATGCAAGCC----- 805
DB 851 GluAlaGluLeuThrLysValArgGluLysHisLeuAlaAlaGluAsnArgLeuThrGlu 870
QY 806 -----CTCATG----- 811
DB 871 MetGluThrMetGlnSerGlnLeuMetAlaGluLysLeuGlnLeuGlnGlnLeuGln 890
QY 812 -----CTTGCCATATGTGAAGGCTCATGAGATAGTGGCATGCTTCTT-----CAG 859
DB 891 AlaLysThrGluLeuCysAlaGluAlaGluGluLeuA-gAlaArgLeuThrAlaLysLys 910
QY 860 CAAATGTTGAGCTCTTGTCTGAAGACATACATGGAATTAATGCGAAGCGTTATGCTGCT 919
DB 911 GlnGluLeuGluGluIleCysHisAspLeuGluAlaA-gValGluGluGluGluArg 930
QY 920 GCTCGTGGAGTTAATTACATTCATCAACACCTTTTGGACATATACGAAATTTACTCTAAA 979
DB 931 CysGlnTyrLeuGlnAlaGluLysLysLysMetGlnGlnAsnIleGlnGlnLeuGluGlu 950
QY 980 AATCCTCAAAATACCAATCCACAGAACATCTACAGGAACACCTGATGAGCGTGCACCC 1039
DB 951 GlnLeuGluGlu-----GluGluSerAlaArgGln 960
QY 1040 TTGGCGGGAAGAACACCTGCACACGGCTGAAAGCTTGTGGAAGAAACACCTGCAGCGCT 1099
DB 961 LysLeuGlnLeuGluLysValThrThrGluAlaLysLeuLysLysLeu-----GluGluAsp 979
QY 1100 GCACGCTGGTGGAGGACGCTGCCAAAATTCATGCTGTGGGAAAGCAACATCTGGA 1159
DB 980 GlnIleIleMetGluAspGlnAsnCysLysLeuAlaLysGluLysLeuLeuGluAsp 999
QY 1160 AGTTTGAACATCAACA-----GAGAAACACCTAGGAAATTTTGAGG 1204
DB 1160 AGTTTGAACATCAACA-----GAGAAACACCTAGGAAATTTTGAGG 1204

```

```

DB 1000 ArgValAlaGluPheThrThrAspLeuMetGluGluGluLysSerLysSerLeuAla 1019
QY 1205 CCTACAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCTAGCAAGATC 1264
DB 1020 LysLeuLysAsnLysHisGluAlaMetIleThrAspLeuGluGluA-gLeuArgArg--- 1038
QY 1265 ACATGGGAGGAAAAAACAACATCTGTAAAGACTGATGATCGCTGGCAGGAGTAACACCTAAT 1324
DB 1039 -----GluGluLysGlnArgGlnGluLys-----ThrArgArg 1051
QY 1325 AAAACTGAAGTTTGGAAAAAGAACATCTAATATGATTGCA----- 1366
DB 1052 LysLeuGluGlyAspSerThrAspLeuSerAspGlnIleAlaGluLeuGlnAlaGlnIle 1071
QY 1367 -----TGTCTACAAAGAAACAATCTACAAAAGCAAGTACAAAT 1405
DB 1072 AlaGluLeuLysMetGlnLeuAlaLysGluGluGluLeuGlnAlaAlaLeuAlaArg 1091
QY 1406 GTGATGTGATGTTCTGTAGAGCCTATATTCACTTTTGGCACACGAGTATTGAAAT 1465
DB 1092 ValGluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGlu--- 1110
QY 1466 TCACAGTGTCAAAAGTTGAGGAAGACTTTAATCTTCTACCAAGATTATCTCTAAGAGT 1525
DB 1111 ThrGlnIleSerGluLeuGlnGluAspLeuGlu-----SerGluArg 1124
QY 1526 GCTGCACAGAAATTATACGTGTTTACCTGATGCTACATATCAAAAAGATATCAAAACAATA 1585
DB 1125 AlaCysArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuAlaLeu 1144
QY 1586 AATCACAATAAGAGATCAGATGTTCCCATCAGATCCAAACGAGAG----- 1633
DB 1145 LysThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGlnLeuLeuArgSerLys 1164
QY 1634 GAAGATGAAGAATATTCTGGGATTTCTGGAGTCTCTTTCAGAGTCTCGAAAGACATCAA 1693
DB 1165 ArgGlnGlnGluValSerIleLeuLysLysThrLeu---GluAspGluAlaLysThrHis 1183
QY 1694 GTGTGTATACGTGATCTATGATCAGAAAGTAATGAGATAAATAGAGAGTAGAAGAG 1753
DB 1184 GluAlaGlnIleGlnGluMetArgGlnLysHisSerGlnAlaValGluGluLeuAlaGlu 1203
QY 1754 CTTCTCAGAAAGCCATCTGCTTCAAGCCTGCGCTGAAATGCAAAAG---ACTGTTCCA 1810
DB 1204 GlnLeuGlnGlnThrLysArgValLysAlaThrLeuGluLysAlaLysGlnThrLeuGlu 1223
QY 1811 AATAAGCCCTTTGAATTGAAGAATGACAAACATGAGCAACCTGAGCAGCTCAGATGTTCCCATCA 1870
DB 1224 AsnGluArgGlyGluLeuAlaAsnGluValLysAlaLeuLeuGlnGlyLysGlyAspSer 1243
QY 1871 GAATCCAAACAAAGGACGATGAAGAAATTTCTGGGATTTCTGAG-----AGTCCC 1921
DB 1244 GluHisLysArgLysLysValGluAlaGlnLeuGlnGlnLeuGlnValLysPheSerGlu 1263
QY 1922 TGTGAGACGGTTTCACAGAGGATGTGATTATTCACCAAGCTACACATCAAAAGAAATTC 1981
DB 1264 GlyCuuArgVal---ArgThrGluLeuAlaAspLysValSerLysLeuGlnValGluLeu 1282
QY 1982 GATACCTTAAGTGGAAAAATTTAGAAAGCTCTCTGTTAAAGATGCTCTTCTGAAAGCCTACC 2041
DB 1283 AspSerValThrGlyLeuLeuAsnGlnSerAspSerLysSerLysLeu----- 1299
QY 2042 TGTGAAGGAAAGTTTCTCTCTTCCAAATTAAGCCTTAGAATTAAGGACAGAGAAACATTC 2101
DB 1300 -----ThrLysAspPheSerAlaLeu 1306
QY 2102 AAGACGAGTCTCCTGATAAAGATGCTTCTGAAGCCTACCTGCTGGAAGGAAAGTTTCT 2161
DB 1307 GluSerGlnLeuGlnAspThrGlnGluLeuLeuGlnGlnGluAsnArgGlnLysLeuSer 1326
QY 2162 CTTCCAAATAAAGCCTTAGAATTAAGGAC-----AGAGAAACACTCAAA 2206
DB 1327 LeuSerThrLysLeuLysGlnMetGluAspGluLysAsnSerPheArgGlnGlnLeuGlu 1346

```

QY 2207 GCAGAGCTCTCTGATGATGATGCTCTCTGAG-----CTTACCTGTGGAGGAAA 2257
 Db 1347 GluGluGluGluGluAlaLysArgAsnLeuGluGluGlnIleAlaThrLeuHisAlaGln 1366
 QY 2258 GTTCTCTCTCCAAATAAAGCTTTAGAA-----TTCAAGGACAGACAAACA 2302
 Db 1367 ValThrAspMetLysLysMetGluAspGlyValGlyCysLeuGluThrAlaGluGlu 1386
 QY 2303 TTCAAGACAGCTCAGATGTTCCCATCAGAA-----TCCAAACAAAGGATGATGAGAA 2356
 Db 1387 AlaLysArgLeuGluGlnLysAspLeuGluGlyLeuSerGlnArgLeuGluGluVal 1406
 QY 2357 AATCTTGGATTTGAGAGTTTCTTGGAGCTCTTACAGAA-----CATGTGTGT 2410
 Db 1407 AlaAlaLysAspLysLeuGluGlyThrLysThrArgLeuGlnGlnLeuAspLeu 1426
 QY 2411 TTACCCAGGCTACACAT-----CAAAAAGAA 2437
 Db 1427 LeuValAspLeuAspHisGlnArgGlnSerValSerAsnLeuGluLysLysGlnLysLys 1446
 QY 2438 TTCGAT-----ACCTAAGTGGAAATTTAGAGAGCTCTCTGAT 2476
 Db 1447 PheAspGlnLeuLeuAlaGluGluLysThrIleSerAlaLysThrAlaGluGluArg 1466
 QY 2477 AAGATGGTCTTCTGAGGCTACCTCTGGATGAAAATTTCTCTCCAAATAAAGCCTTA 2536
 Db 1467 Arg-----AlaGluAla 1470
 QY 2537 GAATTTGAAGGACAGAGAAACATTTCAAAGCAGAGGATGTCAGTTCTGTAGAGTCCACATTC 2596
 Db 1471 GluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGluGluAlaMet 1490
 QY 2597 AGTCTTTTGGCAACCGACTACTGAAATTTACAGTCTTACAAAAGTTGAGAGACATT 2656
 Db 1491 GluGlnLysAlaGluLeuGluArgLeuAsnLysGlnPheArgThrGluMetGluAspLeu 1510
 QY 2657 AATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACTGGCAACAGGAAAGCTGATATT 2716
 Db 1511 --MetSerLysAspValGlyLysSerVal-----HisGluLeu 1524
 QY 2717 GGCATTATTGAAGCTCCACAGATCAACAATAAGATGCCACATCAGAAATTAGA 2776
 Db 1525 GluLysSerAsnArgAlaLeuGluGlnValGluGluMetLysThrGlnLeuGluGlu 1544
 QY 2777 AGAAAGACATACAAATCACTTCAGATTCTGAGATT-----ATCTCTGTGAGTGATACA 2833
 Db 1545 LeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuGluVal-----1561
 QY 2834 CAGAATTATGAGTGTTTACCTGAGGCTACATATCAAAAAGAAATAAAGACACAATGCG 2893
 Db 1562 --AsnLeuGlnAlaMet-----LysAlaGlnPheGluArgAspLeuGln-----Gly 1576
 QY 2894 AAAATAGACAGTCTCTGAAAGCCTTCTCATT-----GAGCCTGCC 2938
 Db 1577 ArgAspGluGlnSerGluGluLysLysGlnLeuValArgGlnValArgGluMetGlu 1596
 QY 2939 ACTGAATGCAAACTCTGTTCCAAATAAAGGCTTAGAATGGAATAAACAACATTG 2998
 Db 1597 AlaGluLeuGluAspGluArgLysGlnArgSerIleAlaMetAlaAlaArgLysLysLeu 1616
 QY 2999 AGACGATTCACACTACCTTATCAAAATCTTGATGCA-----3037
 Db 1617 GluMetAspLysAspLeuGluAlaHisIleAspThrAlaAsnLysAsnArgGluGlu 1636
 QY 3038 -----CTTCTCTTCTTGTGAAGAGGAGGAGAA 3064
 Db 1637 AlaIleLysGlnLeuArgLysLeuGlnAlaGlnMetLysAspCysMetArgAspValasp 1656
 QY 3065 CTTAAAAAGATACTGTGACAAATTTACAGCA-----AAATG 3103
 Db 1657 AspThrArgAlaSerArgGluGluGluIleAlaGlnAlaLysGluAsnGluLysLysLeu 1676

QY 3104 GAACAATGAAAAATAGTTTGTGTACTACAAAAGGAACTGTCAAGCGAAGAAATA 3163
 Db 1677 LysSerMetGluAlaGluMetIleGlnLeuGlnGluLeuAlaAlaGluArgAla 1696
 QY 3164 AAATCAGATTAGAACCAAAAAGCTTAATAGGAAACAGAGCTCTGAGTGTG-----3217
 Db 1697 LysArgGlnAlaGlnGlnArgAspGluLeuAlaAspGluIleAlaAsnSerSerGly 1716
 QY 3218 -----AGATTGCTTTAATCAAGAAGAGAGACAGAGAA-----AATGTCGAT 3262
 Db 1717 LysGlyAlaLeuAlaLeu-----GluGluLysArgArgLeuGluAlaLeuIleAla 1733
 QY 3263 ATATTAAACAAAAATTTAGACCCGAGAGCAACTTAGGAAAAAGTTAGAAAGTGAACAC 3322
 Db 1734 LeuLeuGluGluGluLeuGlu-----GluGluGlnGlyAsnThrGluLeu-----1748
 QY 3323 CAATTTGAACAGACTCTCAGAATACAGATATAGAATTGAAAAGTGTAAACAATAATTG 3382
 Db 1749 --IleAsnAspArgLeuLysLysAlaAsnLeuGlnIleAspGlnIleAsnThrAspLeu 1767
 QY 3383 AAT---CAGCTTTCTCACTCATGAAAGTGAATAAT-----GATCTCTTTTCATGAA 3430
 Db 1768 AsnLeuGluArgSerHisAlaGlnLysAsnGluAsnAlaArgGlnGlnLeuGluArgGln 1787
 QY 3431 AATTGCAATGTTGAAAAAGGAAATTCCTGCTCTAAAAGTGAAGTAGCCACACTGAAACAT 3490
 Db 1788 AsnLysGluLeuLys-----AlaLysLeuGlnGlu 1797
 QY 3491 CAACACAGGTTGAAGGAAAAATAATCTTTGAGGACATTAAAGTTTACAAAGAAAAAGAT 3550
 Db 1798 MetGluSerAlaValLysSerLysLysLysAlaSerIleAlaAlaLeuGluAlaLysIle 1817
 QY 3551 GCTGAATCTCAATGACCTTAAACTGAAACAGAAACAGACATAACAAGGGCATCTCAG 3610
 Db 1818 AlaGlnLeuGluGlnLeuAspAsnGluThrLysGluArgGlnAlaAlaSerLysGln 1837
 QY 3611 TATAGAGCAGCTTAAAGTTCTG 3634
 Db 1838 ValArgArgAlaGluLysLysLeu 1845
 RESULT 15
 MYHB CHICK
 ID MYHB CHICK STANDARD; PRT; 1978 AA.
 AC P10587;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, gizzard smooth muscle.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88118918; PubMed=2892941;
 RA Yanagisawa M., Hamada Y., Katsuragawa Y., Imamura M., Mikawa T.,
 RA Masaki T.;
 RT "Complete primary structure of vertebrate smooth muscle myosin heavy
 RT chain deduced from its complementary DNA sequence. Implications on
 RT topography and function of myosin.";
 RL J. Mol. Biol. 198:143-157(1987).
 RN [2]
 RP REVISIONS.
 RA Masaki T.;
 RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-203.
 RX MEDLINE=88032919; PubMed=3312184;
 RA Maita T., Onishi H., Yajima E., Matsuda G.;
 RT "Amino acid sequence of the amino-terminal 24 kDa fragment of the
 RT heavy chain of chicken gizzard myosin.";
 RL J. Biochem. 102:133-145(1987).

[4] X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
MEDLINE=98412652; PubMed=9741621; Trybus K.M., Cohen C.;
Dominguez R., Freyzon Y., Trybus K.M., Cohen C.;
"Crystal structure of a vertebrate smooth muscle myosin motor domain
and its complex with the essential light chain: visualization of the
pre-power stroke state."; Cell 94:559-571(1998).
RL Cell 94:559-571(1998).
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMW) and 1 heavy meromyosin (HMW). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06546; CAA29793.1; -
DR PIR; S03166; S03166.
DR PDB; 1BR1; 09-SEP-98.
DR PDB; 1BR2; 09-SEP-98.
DR PDB; 1BR4; 09-SEP-98.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001603; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Calmodulin-binding; Multigene family;
KW 3D-structure.
FT INIT MET 0 0 MYOSIN HEAD-LIKE.
FT DOMAIN 1 790 IQ.
FT DOMAIN 791 820 RODLIKE TAIL (S2 AND LMW DOMAINS).
FT DOMAIN 849 1978 COILED COIL (POTENTIAL).
FT NP_BIND 176 183 ATP.
FT DOMAIN 666 688 ACTIN-BINDING.
FT DOMAIN 767 781 ACTIN-BINDING.
FT MOD_RES 1 1 BLOCKED.
FT MOD_RES 127 127 METHYLATION (TRI-) (POTENTIAL).
FT CONFLICT 127 127 MISSING (IN REF. 3).
FT CONFLICT 204 215 KUTISITQGFPS -> RTPASLKVHLPF (IN REF. 1).
FT STRAND 34 37
FT STRAND 43 47
FT STRAND 56 60
FT STRAND 66 69
FT TURN 72 73
FT STRAND 75 76
FT STRAND 80 82
FT HELIX 83 84
FT TURN 83 84

FT STRAND 87 87
FT TURN 88 88
FT HELIX 96 108
FT TURN 109 110
FT STRAND 113 115
FT STRAND 120 123
FT TURN 130 131
FT HELIX 134 139
FT TURN 140 142
FT HELIX 145 147
FT STRAND 152 166
FT TURN 167 167
FT STRAND 170 175
FT TURN 178 179
FT HELIX 182 197
FT HELIX 218 223
FT TURN 224 225
FT HELIX 226 234
FT STRAND 235 236
FT TURN 242 243
FT STRAND 244 245
FT STRAND 248 255
FT TURN 257 258
FT STRAND 261 269
FT HELIX 273 276
FT TURN 277 277
FT TURN 281 282
FT STRAND 286 286
FT HELIX 287 295
FT HELIX 298 301
FT TURN 302 305
FT TURN 309 311
FT TURN 313 314
FT TURN 316 317
FT HELIX 327 341
FT TURN 342 342
FT HELIX 345 363
FT STRAND 367 367
FT STRAND 376 376
FT HELIX 381 390
FT TURN 391 391
FT HELIX 394 402
FT HELIX 419 449
FT STRAND 458 464
FT STRAND 474 474
FT HELIX 476 506
FT TURN 507 508
FT HELIX 521 528
FT HELIX 536 543
FT HELIX 551 561
FT TURN 562 564
FT TURN 566 567
FT STRAND 568 570
FT STRAND 579 584
FT TURN 585 586
FT STRAND 597 592
FT TURN 594 595
FT HELIX 596 601
FT HELIX 606 613
FT TURN 614 614
FT HELIX 618 623
FT TURN 624 624
FT HELIX 658 674
FT TURN 675 675
FT STRAND 677 684
FT TURN 692 693
FT HELIX 697 706
FT TURN 707 708
FT HELIX 709 718
FT STRAND 722 725
FT HELIX 726 733
FT HELIX 734 737
FT TURN 738 739

FT HELIX 748 759
 FT TURN 760 760
 FT HELIX 763 765
 FT STRAND 766 768
 FT STRAND 772 775
 FT TURN 777 778
 FT HELIX 779 787
 SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;

Alignment Scores: Pred. No.: 7.25e-08 Length: 1978
 Score: 300.00 Matches: 228
 Percent Similarity: 38.46% Conservative: 202
 Best Local Similarity: 20.39% Mismatches: 424
 Query Match: 4.65% Indels: 264
 DB: 1 Gaps: 49

US-09-602-362E-26 (1-3673) x MYHB_CHICK (1-1978)

QY	788	GAGTCTAAATGCACGCGCTCATGCTGCCATATGTGAAGGCTCATCAGAGATAGTCGCGC	847	Db	1096	GlulLeuGlnAlaAlaLeuAlaArgLeuGluuAspGluThrSerGlnLysAsnAlaLeu	1115
Db	884	GlulGlnLysHisThrGlnLeu	898	QY	1550	CCTGATGCTACATATCAAAAGAGATATCAAAACATAAATCAACAAATAGAGATCAGATG	1609
QY	848	ATGCTTTCTTCAGCAAAATGTTGACGCTTTGTTGTAAGACATACATGGAATAACTGCAGAA	907	Db	1116	-----LysLysIleArgGluLeuGluSerHisIleSerAspLeuGln	1129
Db	899	GlulLysLeuGlnAlaGluThrGluLeuTyrAlaGlu	912	QY	1610	TTCCCATCAGATCCAAA-----CGAGAGCAAGATCAAGAAATATTCTTCGGATTCT	1660
QY	908	-----CGTTATGCTGCTGCTCGTGGAGTTAATTACATTCATCAACAACCTTTTG	955	Db	1130	GlulAspLeuGluSerGluLysAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeu	1149
Db	913	GluMetArgValArgLeuAlaAlaLysLys	926	QY	1661	GGGAGTCTCTTTGAGAGTCTTCGAAAGACTCAAGTG-----TGTATACCTGAGTCTATG	1714
QY	956	GAACATATACGAAATTAACCTTAAATCTCTCAAAATACCAATCCAGAGAAACATCTACA	1015	Db	1150	SerGluGluLeuGlu-----AlaLeuLysThrGluLeuGluAspThrLeuAspThrThrAla	1168
Db	927	GlulLeuHisGluMetGluAlaArgIleGlu	944	QY	1715	TATCAGAAAGTAAATGGAGATAAATAGAGAAAGTAGAA-----GAGCTT	1756
QY	1016	GGACACCTGATGAGGTGCACCTTGGCGGAAGAACACCTGACAGCGCTGAAGCTTG	1075	Db	1169	ThrGlnGlnGluLeuArgAlaLysArgGluGlnGluValThrValLeuLysArgAlaLeu	1188
Db	945	GlnLeuGlnAlaGluLysLysMetGlnGlnMetLeuAspLeuGluGlnLeu	964	QY	1757	CCTGAGAGCCATCTGCCTTCAAGCCTGCGCTNGAAATGCCAAAGACTGTTCCAAATAAA	1816
QY	1076	CTGGAAGAAACACCTGACGAGGTGCACGCTTGGTGGAGGAACGCTGCCAAAATTCAA	1135	Db	1189	GlulGluThrArgThrHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGln	1208
Db	965	GlulGlu-----GluGluAlaAlaArg	975	QY	1817	GCCTTTCAATTGAAGATGAACAA-----ACATTGAGCAGCTCAGATGTTCCCATCA	1870
QY	1136	TGCTGGGGAAGCAACATCTGGAAGTTTGACAGTCAACAGACAGAAACACCTAGGAA	1195	Db	1209	AlaValGluGluLeuThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAsp	1228
Db	976	LeuGluLysValThrAlaAspGlyLysIleLysLysMetGluAspAsp	991	QY	1871	GAATCCAAACAAAGAGCAGATGAAGAAATCTTGGGATCTTGAGAGTCCCTGTGAGAGC	1930
QY	1196	ATTTGAGGCTCAAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCT	1255	Db	1229	LysThrLysGlnThrLeuGluLysAspAsnAla-----AspLeuAlaAsnGluIleArgSer	1247
Db	992	IleLeu-----IleMetGluAspGlnAsn	999	QY	1931	GTITTCAG-----AGGATGTCTATTACCCAAAGCTACACATCAAAAGAAATTCGAT	1984
QY	1256	AGGAAGATCATCGGAGGAAAGAAACATCTGTAAGACTGAATGCGTGGCAGGAGTA	1315	Db	1248	LeuSerGlnAlaLysGlnAspVal-----GluHisLysLysLys-----	1260
Db	1000	AsnLysLeuThr-----LysGluArgLysLeuGluGluArgValSerAspLeu	1016	QY	1985	ACCTTAAGTGGAAATTAGAAGAGATCTCTCTGTTAAAGATGCTCTCTGGAAGCCTACCTGT	2044
QY	1316	ACACCTTAATAAACTGAAGTTTGGAAAGAGAACATCTAATATGATTGCTCTCTACA	1375	Db	1261	-----LysLeuGlu-----	1263
Db	1017	ThrThrAsnLeuAlaGluGluGluLys-----AlaLysAsnLeuThrLysLeuLysAsn	1035	QY	2045	GGAGGAAAGTTTCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATTCAAA	2104
QY	1376	AAAGAAACATCTCAAAAGCAAGTCAAAATGTGGATGTGAGTCTCTGTA-----	1423	Db	1264	-----ValGlnLeuGlnAspLeuGlnSerLysThrLysSerAspGlyGluArgValArg	1280
Db	1036	LysHisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluLysSerArg	1055	QY	2105	GCAGAGTCTCTGTATAAGATGGTCTTCTGAAGCTACCTGTGGAGAAAGTTTCTCTT	2164
QY	1424	-----GAGCCT	1429	Db	1281	ThrGluLeuAsnGluLysValHisLysLysLeuGlnIleGluValGluAsnValThrSerLeu	1300
Db	1056	GlnGluLeuGluLysIleLysArgLysLeuGluGlyGluSerSerAspLeuHisGluGln	1075	QY	2165	CCAAATAAAGCC-----TTAGAATTA-----AAGGACAGAGAAACACATCAAGCA	2209
QY	1430	ATATTGAGTCTTTTGGCACACGACTATTGAAATTCACAGTGTACAAAAGTTGAGGAA	1489	Db	1301	LeuAsnGluAlaGluSerLysAsnIleLysLeuThrLysAspValAlaThrLeuGlySer	1320
Db	1076	IleAlaGluLeuGlnAlaGlnIleAlaGluLeuLysAlaGlnLeuAlaLysLysGluGlu	1095	QY	2210	GAGTCTCTGTATATGATGCTCTTCTGAAGCTACCTGTGGAAGAAAGTTTCTCTTCCA	2269
QY	1490	GACTTTAATCTGCTACCAAGATTATCTTAAGAGTCTGCACAGAAATTATACGTTTAA	1549	Db	1321	GlnLeuGlnAspThrGlnGluLeuLeuGlnGluThrArgGlnLysLeuAsnValThr	1340
Db	1095	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1105	QY	2270	AATAAGCTTTAGAAATTAAGGACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCA	2329
QY	1549	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1559	Db	1341	ThrLysLeuArgGlnLeuGluAspAspLysAsnSerLeuGlnGlnLeuAspGluGlu	1360
Db	1095	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1105	QY	2330	-----GAATCCAAACAA-----	2347
QY	1549	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1559	Db	1361	ValGluAlaLysGlnAsnLeuGluArgHisIleSerThrLeuThrIleGlnLeuSerAsp	1380
Db	1095	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1105	QY	2348	GATGAAGAAAATTTCTGGGATTTTTCAGAGATTTCTTCAGAGCTCTCTTACAGAAATGATGTG	2407
QY	1549	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1559	Db	1391	SerLysLysLysLysLeuGlnGluPheThrAlaThrValGluThrMetGluGluGly-----	1398
Db	1095	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1105	QY	2408	TGTTTACCAAGGCTACACATCAAAAGAAATTCATACCTTAAGTGGAAATTAGAGAG	2467
QY	1549	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1559	Db	1399	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1415
Db	1095	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1105	QY	2468	---TCTCTGTATAAAGATGGTCTTCTGAAGCTACCTGTGGAATGAAATTTCTCTTCCA	2524
QY	1549	-----LysLysLysLysLeuGlnArgGluIleGluSerLeuThrGlnGlnPheGluGlu	1559	Db	1416	LysAlaAlaSerTyrAspLysLeuGluLysThrLysAsnArgLeuGlnGlnLeuAsp	1435

QY 2525 AATAAGCCTTAGAATTGAAGGACAGACAGAGAAACATTCAAAGCAGAGGATGTGATCTGTGTA 2584
Db 1436 AspLeuValValAspLeuAspAsnGlnArgGlnLeu- 1451
QY 2585 GAGTCCACATTCAGTCTTTTGGCAACCGACTACTCTGAAAATTCACAG---TCTCAAAA 2641
Db 1452 GluLysLysGlnLysLysPheAspGlnMetLeuAlaGluGluLysAsnIleSerSerLys 1471
QY 2642 GTTGAGGAAGAC-----TTTAATCTTACTACCAAGGAGGAGCAACAAGACAGTA 2692
Db 1472 TyrAlaAspGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeu 1491
QY 2693 ACTGGCAACAGCAAGCTGATATTGGCATTATTGAACGAGCTCCACAGATCAACAAT 2752
Db 1492 SerLeuAlaArgAlaLeuGluAlaLeuGluAlaLysGluGluLeuGluArgThrAsn 1511
QY 2753 AGATGCCACATCAGATTAGGAAGAAAGAGATACAAAATCAACTTCAGAT----- 2806
Db 1512 LysMetLeuLysAlaGluMet-----GluAspLeuValSerSerLysAspVal 1528
QY 2806 ----- 2806
Db 1529 GlyLysAsnValHisGluLeuGluLysSerLysArgThrLeuGluGlnGlnValGluGlu 1548
QY 2807 -----TCTGAGATTATCTCTGTGAGTGATACACAG--- 2836
Db 1549 MetLysThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaAlaGluAspAlaLysLeu 1568
QY 2837 -----AATTATGAGTGTTCCTGAGGCTACATATCAAAAGAAATAAAGACA 2884
Db 1569 ArgLeuGluValAsnMetGlnAlaMet---LysSerGlnPheGluArgAspLeuGlnAla 1587
QY 2885 ACAATGGCAAAATAGCAAGTCTCTGAA-----AAGCCTTCTCATTTCAGCCT 2935
Db 1588 ArgAspGluGlnAsnGluGluLysArgGlnLeuLysGlnLeuHis---GluHis 1606
QY 2936 GCCACTGAATCAAACTCTGTTCCAAATAAGGCTTAGAATGAAGAATAAACAACA 2995
Db 1607 GluThrGluLeuGluAspGluArgLysGlnArgAlaLeuAlaAlaLysLysLys 1626
QY 2996 TTGAGAGCAGATTCAACTACCTATCAAAATCTGGATGCACTCTCTTCTGTCAAAGA 3055
Db 1627 LeuGluValAspValLysAspLeuGluSerGlnValAsp-----SerAlaAsnLys 1643
QY 3056 GGAAGGGAACCTTAAAGAGATAACTGTGAACAAATTCACGCAAAAATGCAACAATGAAA 3115
Db 1644 AlaArgGlu-----GluAlaIleLysGlnLeuArgLysLeuGlnAlaGlnMetLys 1660
QY 3116 AATAAGTTTGTGTACTACAAAGAACTGTCTCAGAGCGAAAGAAATAAATCAGTTA 3175
Db 1661 Asp-----TyrGlnArgAspLeuAspAlaArgAlaAlaArgGluGluIle 1676
QY 3176 -----GAGAACCAAAA-----GCTAAATGGAAACAAGAGCTCTGCAGT 3214
Db 1677 PheAlaThrAlaArgGluAsnGluLysLysAlaLysAsnLeuGluAlaGluLeuIleGln 1696
QY 3215 GTGAGATTCCTTTAAATCAAGAGAGAGAGAGAAATGTCTGATATATTAAGAA 3274
Db 1697 LeuGlnGluAspLeuAlaAlaGluArgAlaArgLysGlnAlaAspLeuGluLysGlu 1716
QY 3275 AAAATTAGACCCGAA-----GAGCACTTAGG 3301
Db 1717 GluMetAlaGluGluLeuAlaSerAlaAsnSerGlyArgThrSerLeuGlnAspGluLys 1736
QY 3302 AAAAAGTTAGAGTGAAA---CACCACTTGAACAGACTCTCAGAAATACAAGATAGAA 3358
Db 1737 ArgArgLeuAlaArgIleAlaGlnLeuGluGluLeuAspGluGluHisSerAsn 1756
QY 3359 TTGAAAAGTGTAACAGTAATTTGATCAG---GTTTCTCAGCTCATGAAGTGAAT 3415
Db 1757 IleGluThrMetSerAspArgMetArgLysAlaValGlnGlnAlaGluGlnLeuAsnAsn 1776

QY 3416 GATCTCTTTCATGAAAATTGCAATCTTGAAGAAAGAA----- 3451
Db 1777 GluLeuAlaThrGluArgAlaThrAlaGlnLysAsnGluAsnAlaArgGlnGlnLeuGlu 1796
QY 3451 ----- 3451
Db 1797 ArgGlnAsnLysGluLeuArgSerLysLeuGlnGlnMetGluGlyAlaValLysSerLys 1816
QY 3452 -----ATTGCCATGCTTAAACTGGAAGTACCCACACTGCAACATCAACACAG 3499
Db 1817 PheLysSerThrIleAlaAlaLeuGluAlaLysIleAlaSerLeuGluGlnLeuGlu 1836
QY 3500 GTGAAGCAAAAATAAATACTTTTGGACATTAAGATTTTCAAGAAAAAGAAATGCTGAACCTT 3559
Db 1837 GlnGluAlaArgGluLysGlnAlaAlaAlaLysThrLeuArgGlnLysAspLysLysLeu 1856
QY 3560 CAATGACCCCTAAACTGAAACAGAAACAGTAAACAAAAAGGCGCATCTCAGTATAGAGAG 3619
Db 1857 LysAspAlaLeu---LeuGlnValGluAspGluArgLysGlnAlaGluGlnTyrLysAsp 1875
QY 3620 CAGCTTAAAGTTCTGACGGCGAGACACCATGCTGACTTCTAAATTTGAAGGAA 3673
Db 1876 GlnAlaGluLysGlyAsnLeuArgLeuLysGlnLeuLysArgGlnLeuGluGlu 1893

Search completed: July 15, 2004, 09:13:01
Job time : 167.051 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 09:09:52 ; Search time 53.6846 Seconds

(without alignments)
13162.486 Million cell updates/sec

Title: US-09-602-362E-26

Perfect score: 6448
Sequence: 1 caagagcttgcatcacaga.....tgacttctaattgaaggaa 3673

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2.1/USPTC spool_p/US0602362/runat_15072004_093626_22015/app_query.fasta.1.10325
-DB=PIR_78 -OFFT=fastan -SUFFIX=pr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0602362@cgn_1_1_248 @runat_15072004_093626_22015 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	347.5	5.4	1427	2 S22695	restin - human
2	343	5.3	1392	2 A43336	microtubule-vesicl
3	342	5.3	1790	2 S67593	transport protein
4	330	5.1	2663	1 S28261	centromere protein
5	324.5	5.0	1938	2 JC5421	smooth muscle myos
6	324.5	5.0	1972	2 JCS420	smooth muscle myos
7	319.5	5.0	2954	2 T14156	kinesin-related pr
8	318.5	4.9	2116	2 A26655	myosin heavy chain
9	315	4.9	1972	1 A41604	myosin heavy chain
10	315	4.9	3488	2 T34418	myosin heavy chain
11	312.5	4.8	5327	2 T13564	hypothetical prote
12	311	4.8	3259	1 A56539	microtubule-associ
13	310	4.8	1981	1 A61231	giantin - human
14	310	4.8	3225	2 I52300	myosin heavy chain
					giantin - human

15	302.5	4.7	3187	2 JC5837	364K Golgi complex
16	302	4.7	1004	2 A55142	myosin-light-chain
17	300.5	4.7	1313	2 F96673	hypothetical prote
18	300.5	4.7	2253	2 T30336	nuclear/mitotic ap
19	300	4.7	1979	1 S03166	myosin heavy chain
20	299.5	4.6	2677	2 A38194	desmoplakin 1 - hu
21	298	4.6	2057	2 S61477	myosin II heavy ch
22	297	4.6	2017	1 A36014	myosin heavy chain
23	294.5	4.6	1538	2 T29095	cardiac muscle fac
24	294	4.6	1939	2 T18372	repeat organellar
25	293.5	4.6	1690	2 T13030	microtubule bindin
26	293.5	4.5	1957	2 T38077	hypothetical coile
27	293	4.5	6642	2 T29757	protein UNC-89 - C
28	292.5	4.5	1410	1 A57013	early endosome ant
29	291.5	4.5	2331	2 T25410	hypothetical prote
30	290	4.5	2139	2 T18296	myosin heavy chain
31	289	4.5	1875	2 S38173	myosin-like protei
32	287.5	4.5	1738	2 T14867	interaptin - slime
33	286.5	4.4	1999	1 S21801	myosin heavy chain
34	286	4.4	993	2 S49461	synaptosomal compl
35	284.5	4.4	1558	2 E71603	RESA-H3 antigen Pf
36	284.5	4.4	1727	2 T50073	myosin-like coiled
37	284	4.4	1642	2 T08880	NMDA receptor-bind
38	281	4.4	1390	2 S51364	sperm tail-specifi
39	280	4.3	1780	2 T17272	hypothetical prote
40	279	4.3	1295	2 T24587	hypothetical prote
41	279	4.3	1959	1 A33977	myosin heavy chain
42	277	4.3	1837	2 T41023	probable nuclear p
43	276.5	4.3	1356	2 S32763	kinectin 1 - human
44	276	4.3	1269	2 F84730	probable myosin he
45	275.5	4.3	1992	2 A47297	myosin heavy chain

ALIGNMENTS

RESULT 1

S22695
restin - human

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999

C:Accession: S22695; S19853

R:BiIbe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;

EMBO J. 11, 2103-2113, 1992

A>Title: Restin: a novel intermediate filament-associated protein highly expressed in th

A:Reference number: S22695; MUID:92289675; PMID:1600942

A:Accession: S22695

A:Molecule type: mRNA

A:Residues: 1-1427 <BIL>

A:Cross-references: EMBL:X64838; NID:g35998; PIDN:CAA46050.1; PID:g35999

C:Keywords: cytoskeleton

Alignment Scores:	7,53e-12	Length:	1427
Pred. No.:	347.50	Matches:	304
Score:	36.79%	Conservative:	207
Percent Similarity:	21.89%	Mismatches:	510
Best Local Similarity:	5.39%	Indels:	368
Query Match:	2	Gaps:	67
DB:			

US-09-602-362E-26 (1-3673) x S22695 (1-1427)

QY	1	CAAGAGCTTGCGGATACAGAAATTTCTGCTGTTGGGGCGGT-----	45
DB	75	GlnPheLeuGlyGluThrGlnPheAlaProGlyGlnTrpAlaGlyTleValLeuAspGlu	94
QY	46	---GCGGAACTGAAGCGGCGGAGTCGAGCCCGG-----	78
DB	95	ProtleGlyLysAsnAspGlySerVal---AlaGlyValArgTyrPheGlnCysGluPro	113
QY	79	---GCGCGGTCTCGGGAAGCGGTAAAGCGGGAAGCGAGCGGAGGGGTAGGGCTGGGAA	135
DB	114	LeuLysGlyIlePheThrArgProSer---LysLeuThrArgLysValGlnAla---Glu	131

Qy	136	GGGCGAGCGGAGCGCGGGCTCTCTTAGCAGGGGGCTGCAGCCATGAAGAGGCTCTTA	195
Db	132	AspGluAlaAenGlyLeuGlnThrThrProAlaSerArgAlaThrSerProLeuCysThr	151
Qy	196	GCTGCGCTGCGCAAGGGGCTGCGGGCCCGAGGCCCGCAAC-----	237
Db	152	SerThrAlaSerMetValSerSerProSerThrProSerAsnIleProGlnIysPro	171
Qy	238	-----CCCTTCAGCAACGGGTCTACACT	261
Db	172	SerGlnProAlaAlaLysGluProSerAlaThrProProIleSerAsnLeuThrLysThr	191
Qy	262	GAGAAGGACTACGGGACCACCTACTCTCGGGATCTAGGAGATGCCATACAGCTGCCTCC	321
Db	192	AlaSerGlu-----SerIleSerAsnLeuSerGlu	201
Qy	322	CGGGGCCCAAGTCCAGAACTGGAGAAGATGACAGTAGGGAAGAAGCCGCTCAACCTGAAC	381
Db	202	AlaGlySerIleLysLys-----GlyGluArgGluLeuLysIleGly	215
Qy	382	AAAGAGAGATATCAAGAGAGACTGCTCTACACTGGGCTGTGTCAATGCC---CATGCA	438
Db	216	AspArg-----ValLeuValGlyGlyThrLysAla	225
Qy	439	NAAGTAGTAACATTCTGGTAGACAGAAAGTCCNGCTTAATGTCCTGATGGCGAAGGG	498
Db	226	GlyValValArgPheLeu-----Gly	232
Qy	499	AGGACACTCTGATGAAGGCTCTCAATGC-----	528
Db	233	GluThrAspPheAlaLysGlyGluTrpCysGlyValGluLeuAspGluProLeuGlyLys	252
Qy	529	-----GAGAGGCAAGCTTTGTGCAATATCTCATAGATGCTGG	567
Db	253	AsnAspGlyAlaValAlaGlyThrArgTyrPheGlnCysGlnProLysTyrGlyLeuPhe	272
Qy	568	TGCTGATCTAAATATATAGTAGTGTATGGCAACACGGCTCTCCATATGCGGTTTATAG	627
Db	273	AlaProValHisLysValThrLysIleGlyPheProSerThrThrProAlaLysAlaLys	292
Qy	628	TGAGAAATTA-----TTAATGTGGCAACACTGCTGCTTATGGTCAGTCATCGA	678
Db	292	sAlaAsnAlaValArgArgValMetAlaThrSerAlaSerLeuLysArgSerProSe	312
Qy	679	GGTCAAAACAAAGCTAGCTCACACCCCTTTTACTGGCCATACAGAAAAGAACAGCA	738
Db	312	rAlaSerSerLeuSerSerMetSerValAlaSerSerValSerArgProSerAr	332
Qy	739	AACGTGGAATTTTACTACAAAATATGCAAAATGCAACGCATTTAATGAGTCTAAATG	798
Db	332	gThr-----GlyLeuLeuThrGluThrSerSerArgTyrAlaArgLysIleSerGlyTh	350
Qy	799	CACAGCCTCATGCTGCTCATATGTGAAGGCTCATCAGAGACTGCGCATGCTTCTTCA	858
Db	350	rThrAlaLeuGlnGluAlaLeuLysGluLysGlnGlnHisIleGlnLeuLeuAlaG	370
Qy	859	GCAAAATTTTGACGCTCTTGTCTGAAGACATACATGGAATAACTCCAGAACGTTATGCTGC	918
Db	370	uArgAspLeuGlu-----ArgAlaGluValAlaLysAla	381
Qy	919	TGCTCGTGGAGTTAATTACATTCATCAACACTTTTG-----GA	957
Db	381	aThrSerHisValGlyGluIleGluGlnGlnLeuAlaLeuAlaArgAspGlyHisAspG	401
Qy	958	ACATATACGAAAAATTACTCTAAAAATCTCTCAAAATACCAATCCAGAGGAACATCTACAGG	1017
Db	401	nHisValLeuGluLeu-----GluAla	408
Qy	1018	AACACCTGATCAGCGCTGCACCCCTTGGCGGAAGAAACACCTGCACGCGTGAACGTTGCT	1077
Db	408	aLysMetAspGlnIleuArgThrMetValGluAlaAlaAspArgGlnLysValGluLeu	428
Qy	1078	GGAAAAACACCTCAGCAGGGCTGCACGCTGGTGGAGGGAACGCTGCCAAAAATTCATG	1137

Db	428	uAsnGlnLeu--	GlUGluGluLysArgLysValGluAspLeuGlnPheArgValGlu--	446
Qy	1138	TCTGGGAAAGCAACATCTGGAAAGTTTGAACAGTCAACAGCAAGAAACACCTCTGGAAAT	1197	
Db	447	GluGluSerIleThrLysGlyAspLeuGluThrGlnThrLysLeuGluHisAlaArgGIl	466	
Qy	1198	TTTGAGGCCTACAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCTAG	1257	
Db	466	e-----LysGluLeuGluGlnSerLeuLeuPhe---GluLysThrLysAlaAs	481	
Qy	1258	GAAGATCATGTGGAGGAGAAAGAAACATCTGTAAACAGCTGAATGCTGCGCAGGAGTAAC	1317	
Db	481	plysLeuGlnArgGluLeuGluAspThrArgValAlaThr-----Va	495	
Qy	1318	ACCTAATAAACTGAAGTTTGGAAAAAGAAACATCTAATAATGATGTCATGCTCTACAAA	1377	
Db	495	lserGluLysSerArgIleMetGlu-----LeuGluLysAspLeuAlaLeuArgValGIl	513	
Qy	1378	AGAAACA-----TCTCAAAAGCAGAGTACAAATGTGGAT--	1411	
Db	513	ngluValAlaGluLeuArgArgLeuGluSerAsnLysProAlaGlyAspValaspMe	533	
Qy	1412	----GTGAGTTCTGTAGAGCCTATATCACTCTTTTGGC-----ACACGGAC	1455	
Db	533	tSerLeuSerLeuLeuGlnLilieSerSerLeuLeuGlnLysLeuGluValThrArgTh	553	
Qy	1456	TATTGAAATTCACAGTGTACAAAGTTTGGAGAGACTTTAATCTTGCTACCAAGATTAT	1515	
Db	553	rAspHisGlnArgGluIleThrSerLeuLysGluH:spHeGlyAlaArg-----	569	
Qy	1516	CTCTAAGAGTGTCTGCACAGAATTATACGTCTTTACTGTGATCTCATATCAAAAGAGATAT	1575	
Db	570	-----GluGluThrHisGlnLysGluIl	577	
Qy	1576	CAAA-----ACAATAATCAAAAATAGAGATCAGATGTTCCCATCAGAAATCCAA	1626	
Db	577	eLysAlaLeuThrAlaThrGluLysSerLysGluAsnGluSerLeuLysSerLy	597	
Qy	1627	ACGAGAGGAAGATCAAGAATATTCTTGGGATCTCGGAGTCTCTTTGAGAGTTCTGCAAA	1686	
Db	597	sLeuGluHisAlaAsnLysGluAsnSerAspValIleAlaLeuTrpLysSerLysLeuGIl	617	
Qy	1687	GACTCAAGTGTATACCTGAGTCTATGTATCAGAAAGTAAATGGAG-----	1732	
Db	617	uThrAlaIle-----AlaSerHisGlnAlaMetGluGluLeuLysValSe	633	
Qy	1733	-----ATAATAGAGAGTAGAAGAGCTTCCTGAG---AAGCCATCTGCCTT	1776	
Db	633	rPheSerLysGlyLeuGlyThrGluThrAlaGluPheAlaGluLeuLysThrGlnIlleGl	653	
Qy	1777	CAAGCCTCCCGTNGAAATGCAAAAGACTGTTCCAAT-----	1813	
Db	653	uLysMetArgLeuAspTyrGlnHisGluIlleGluAsnLeuGlnAsnGlnGlnAspSerGl	673	
Qy	1814	-AAAGCCTTGAATGGAAGATGACAAACATGTAGAGCGAGCTCAGATGTTCCCATCAGA	1872	
Db	673	uArgAlaAlaHisAlaLysGluMetGluAlaLeuArgAlaLysLeuMet-----	699	
Qy	1873	ATCCAAACAAAGGACCATCAAGAAATTTCTGGGATTTCTGAGAGTCCCTGTGACAGCGGT	1932	
Db	690	----LysValIleLysGluLysGluAsnSer-----LeuGluAlaIl	702	
Qy	1933	TTACAGAGAGATGTGTATTATCCCAAGCT-----ACATCAAAAGAAATTCAGATC	1986	
Db	702	eArgSerLys-----LeuAspLysAlaGluAspGlnHisLeuValGluMetGluAs	719	
Qy	1987	CTTAAGTCGAAATTAGAAGAGTCTCCTGTTAAAGAT-----GGTCTTCTGAAGCC	2037	
Db	719	pThrLeuAsnLysLeuGlnGluAlaGluIleLysValLysGluLeuGluValleuGlnAl	739	
Qy	2038	TACCTGTGGAGGAAAGTTTCT---CTTCCAATAAGCCCTTGAATTAAGGACAGAGA	2094	

Db 739 aLysCysAsnGluGlnThrLysValIleAspAsnPheThrSerGlnLeuLysAlaThrG1 759
QY 2095 AACATTCAAAGCAGAGTCTCTGTGATAAAGAGTGGTCTTCTGAAAGCCTCACTCTGGAAGGAA 2154
Db 759 u-----GluLysLeuLeuAspLeuAspAlaLeuArgLy 770
QY 2155 AGTTTCTCTTCCAAATAAGCCTTAGAATTAAAGCAGACAGAAACACTCAA----- 2206
Db 770 sAlaSerSerGluGlyLysSerGluMetLysLysLeuArgGlnGlnLeuGluAlaG1 790
QY 2207 -----CCAGAGTCTCTCTGATAATGATGGTCT 2232
Db 790 uLysGlnIleLysHisLeuGluIleGluLysAsnAlaGluSerSerLysAlaSerSerI1 810
QY 2233 TCTGAAGCCTACTCTGTGAAGAAAGTTTCTCTTCCAAATAAGCTTTAGAAATTAAGGA 2292
Db 810 eThrArgGluLeuGlnGlyArgGluLeuLysLeuThrAsnLeuGlnGluAsnLeuSerG1 830
QY 2293 C-----AGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACA 2340
Db 830 uValSerGlnValLysGluThrLeuGluLysGluLeuGlnIleLeuLysGluLysPheAl 850
QY 2341 AAAGGATGATGAAGAAATTTCTTGGGATTTTGAGATTTCTCTGAGACTCTCTTACAGAA 2400
Db 850 aGluAlaSerGluGluAlaValSerValGlnArgSerMetGlnGluThrVal----- 867
QY 2401 TGATGTGTCTTACCAAGCTACACATCAAAAGAA-----TTGCATACCTTAAGTGG 2454
Db 868 -----AsnLysLeuHisGlnLysGluGlnPheAsnMetLeuSerSe 882
QY 2455 AAAATTAAGAGTCTCTGTGATAAAGTGTCTTCTGAAGCCTACTGTGGAATGAAAT 2514
Db 882 rAspLeuGluLys-----LeuArgGluAsnLeuAlaAspMetGluAl 896
QY 2515 TTCTCTTCCAAATAAGCCTTAGAATTGAAGCAGAGAAACATTCAAGCGAGGATGT 2574
Db 896 aLysPheArgGluLys-----AspGluArgGluGlnLeuIleLysAlaLysGlu-- 913
QY 2575 GAGTCTGTAGAGTCCACATTCAGTCTTTTGGCAACCGACTACTGAAAT-----TCACA 2631
Db 914 ----LysLeuGluAsnAspIleAlaGluIleMetLysMetSerGlyAspAsnSerSerG1 932
QY 2632 GTCTCAAAAGTTGAGGAGATTTAACTTACTTACCAAGGAGGAGCAACAAGACAGT 2691
Db 932 nLeuThrLysMetAsnAspGluLeuArgLeu----- 942
QY 2692 AACTGCACACAGGACGTGATTT----- 2716
Db 943 -----LysGluArgAspValGluGluLeuGlnLeuLysLeuThrLysAlaAsnG1 959
QY 2717 -----GGCATTATTGAACGAGCTCCACAAGATCAACAATAAGATGCCACATCA-- 2767
Db 959 uAsnAlaSerPheLeuGlnLysSerIleGluAspMetThrValLysAlaGluGlnSerG1 979
QY 2768 -----GAATTAGAGAAAGAAAGATACAAAATCAACTTCAGA 2805
Db 979 nGlnGluAlaLysHisGluLysHisGluLysGluLeuGluArgLysLeuSerAs 999
QY 2806 TTCTGAGATTATCTGTGAGTGATACACAGAAATTATGAGTCTTTACCT---GAGGCTAC 2862
Db 999 pLeuGlu-----LysLysMetGluThrSerHisAsnGlnCysGlnGluLeuLysAlaAr 1017
QY 2863 ATATCAAAAA-----GAATAAAGACACAATAATGCAAAATAGAAAGATCTCTGGA 2913
Db 1017 gTyrGluArgAlaThrSerGluThrLysThrLysHisGluGluIleLeuGlnAsnLeuG1 1037
QY 2914 AAG-----CCTTCTCATTGAGCTGCCACTGAAATCAAACTCTCT 2958
Db 1037 nLysThrLeuLeuAspThrGluAspLysLysGlyAlaArgGlu---GluAsnSerG1 1056
QY 2959 TCCAAATAAGGCTTAGAATTGAAGATAAACAACATTGAGAGCAGATTCACCTACCT 3018
Db 1056 yLeuLeuGlnGluLeuGluLeuArgLysGlnAlaAspLysAlaLysAlaAlaGlnTh 1076

QY 3019 ATCAAAAATTTGGATGCATCTCTCTTCTGAAAAGAGGAACTTAAAAAGATAA 3078
Db 1076 rAlaGlu-----AspAlaMetGlnIleMetGlu-----GlnMetThrLysGluLy 1091
QY 3079 CTGTGAACAAATTACAGAAAATGGAACAAATGAAAATAAGTTTGTGTACTACAAA 3138
Db 1091 sThrGluThrLeu---AlaSerLeuGluAspThrLysGlnThrAsnAlaLysLeuGlnAs 1110
QY 3139 GGAACCTGTGAGACCGAAGAA-----ATPAAATCACAGTTAGAGAACCAAAAAGCTAA 3192
Db 1110 nGluLeuAspThrLeuLysGluAsnLeuLysAsnValGluGluLeuAsnLysSerLy 1130
QY 3193 ATGGRAACAGAGCTGTGAGTGTGAGATTCCTTTAAATCAAGAA---CAAGAGAAGAG 3249
Db 1130 s-----GluLeuLeuThrValGlu-----AsnGlnLysMetGluGluPheAr 1144
QY 3250 AAGAAATGTCGATATATTAATA----- 3271
Db 1144 gLysGluIleGluThrLeuLysGlnAlaAlaGlnLysSerGlnGlnLeuSerAlaLe 1164
QY 3272 -----GAAAAATTAGACCCGAGCAACTTAGAAAAAG----- 3307
Db 1164 uGlnGluGluAsnValLysLeuAlaGluGluLeuGlyArgSerArgAspGluValThrSe 1184
QY 3308 -----TTAGAAGTGAACA 3321
Db 1184 rHisGlnLysLeuGluGluArgSerValLeuAsnAsnGlnLeuLeuGluMetLysLy 1204
QY 3322 CCACCTTGAACAGACTCTCAGATACAGATATAGATTAAGAAAGTACACAGTAATT 3381
Db 1204 sArgGluSerLysPheIleLysAspAlaAspGluGluLysAlaSerLeuGlnLysSerIl 1224
QY 3382 GAATCAGGTTTCTCACACTCATGAAAGTGAATGATCTCTTTTCATGAAAAATTCATGTT 3441
Db 1224 eSer---IleThrSerAlaLeuThrGluLysAspAlaGluLeuGluLys-----Le 1241
QY 3442 GAAAGAGAAATTCCTGATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 3501
Db 1241 uArgAsnGluValThrValLeuArgGlyGluAsnAlaSerAlaLysSerLeuHisSerVa 1261
QY 3502 GAAGAAAATAATATCTTTGAGGACATTAAGATT---TTACAAGAAAAGAAATCTGAAC 3558
Db 1261 lValGlnThrLeuGluSerAspLysValLysLeuGluLeuLysValLysAsnLeuGluLe 1281
QY 3559 TCAATGACCTTAAACTGAAACAG 3583
Db 1281 uGlnLeuLysGluAsnLysArgGln 1289

RESULT 2

A43336
microtubule-vesicle linker CLIP-170 - human
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C:Accession: A43336
R:Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992

A:Title: CLIP-170 links endocytic vesicles to microtubules.
A:Reference number: A43336; MUID:92405160; PMID:1956075
A:Accession: A43336
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1392 <PIE>
A:Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622

Alignment Scores:

Pred. No.:	1.33e-11	Length:	1392
Score:	343.00	Matches:	306
Percent Similarity:	35.95%	Conservative:	189
Best Local Similarity:	22.22%	Mismatches:	503
Query Match:	5.32%	Indels:	379
DB:	2	Gaps:	69

QY 1909 TTCTGAGAGTCCCTGTGAGACGGTTTCACAGAGGATGTGATTATCCCAAGACT- 1963
 Db : : : : :
 QY 664 -----LeuGluAlaileArgSerLys-----LeuAspLysAlaGluAs 676
 QY 1964 -ACACATCAAAAAGATTCCGATACCTTAAGTCGAAAAATTAGAAAGTCTCCTGTTAAAGA 2022
 Db : : : : :
 QY 676 pGlnHisLeuValGluMetGluAspThrLeuAsnLysLeuGlnGluAlaGluileLysVa 696
 QY 2023 T-----GCTCTTCTCAAGCCTACCTGTGGAAGAAAGTTTCT--CTTCCAAATAA 2070
 Db : : : : :
 QY 696 lLysGluLeuGluValLeuGlnAlaLysCysAsnGluGlnThrLysValIleAspAsnPh 716
 QY 2071 AGCTTAGAATTAAAGGACAGAGAAACATTCAAGACAGAGTCTCTGATGAAGATGGTCT 2130
 Db : : : : :
 QY 716 eThrSerGlnLeuLysAlaThrGlu-----GluLysLe 727
 QY 2131 TCTGAAGCCTACCTGTGGGAAGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAAGGA 2190
 Db : : : : :
 QY 727 uLeuAspLeuAspAlaLeuArgLysAlaSerSerGluGlyLysSerGluMetLysLysLe 747
 QY 2191 CAGAGAAACACTCAAA-----GC 2208
 Db : : : : :
 QY 747 uArgGlnGlnLeuGluAlaAlaGluLysGlnIleLysHisLeuGluileGluLysAsnAl 767
 QY 2209 AGAGTCTCTGATATGATGCTTCTGAAGCCTACCTGTGGAAGAAAGTTTCTCTTCC 2268
 Db : : : : :
 QY 767 aGluSerSerLysAlaSerSerIleThrArgGluLeuGlnGlyArgGluLeuLysLeuTh 787
 QY 2269 AAATAAAGCTTTAGAAAGGAC-----AGAGAAACATTCAAGACAGCTCA 2316
 Db : : : : :
 QY 787 rAsnLeuGlnGluAsnLeuSerGluValSerGlnValLysGlnThrLeuGluLysGluLe 807
 QY 2317 GATGTTCCCATCAGATCCAAACAAAGGATGATGAAGAAATTTCTGGATTTGAGAG 2376
 Db : : : : :
 QY 807 uGlnIleLeuLysGluLysPheAlaGluAlaSerGluGluAlaValSerValGlnArgSe 827
 QY 2377 TTCTCTGAGACTCTCTACAGAATGATGTGTGTACCAAGGCTACACATCAAAAGA 2436
 Db : : : : :
 QY 827 rMetGlnGluThrVal-----AsnLysLeuHisGlnLysG 839
 QY 2437 A-----TTCGATACCTTAAGTGGAAATTTAGAGAGTCTCTGATGAAGATGGTCTCT 2490
 Db : : : : :
 QY 839 uGluGlnPheAsnMetLeuSerSerAspLeuLys-----LeuAr 853
 QY 2491 GAAGCCTACCTGTGAATGAAATTTCTTCTCAATAAAGCCTTAGAATTGAAGGACAG 2550
 Db : : : : :
 QY 853 gGluAsnLeuAlaAspMetGluAlaLysPheArgGluLys-----AspGluArgGluG 871
 QY 2551 AGAAACATTCAAGCAGAGGATGTGAGTTCTGTAGAGTCCACATTCACTTTTGGCAA 2610
 Db : : : : :
 QY 871 uGlnLeuileLysAlaLysGlu-----LysLeuGluAsnAspIleAlaGluileMetLy 889
 QY 2611 ACCGACTACTGAAT--TCACAGTCTACAAAGTTGAGGAGAGTCTTAATCTTACTAC 2667
 Db : : : : :
 QY 889 sMetSerGlyAspAsnSerSerGlnLeuThrLysMetAsnAspGluLeuArgLeu----- 907
 QY 2668 CAAGGAGGAGCAACAAACAGACTAAGTGGACACAGGAACCTGATATTT----- 2716
 Db : : : : :
 QY 908 -----LysGluArgAspValGluGluLeuG 916
 QY 2717 -----GGCATATTGAACAGGCTCCACAGATCA 2745
 Db : : : : :
 QY 916 nLeuLysLeuThrLysAlaAsnGluAsnAlaSerPheLeuGlnLysSerIleGluAspMe 936
 QY 2746 AACAAATAAGATGCCACATCA-----GAATTAGGAAGAAA 2781
 Db : : : : :
 QY 936 tThrValLysAlaGluGlnSerGlnGlnGluAlaLysLysHisGluGluGluLysL 956
 QY 2782 AGAAGATACAAATCAACTTCGATCTGAGATTATCTCTGTGAGTACATACAGAAATTA 2841
 Db : : : : :
 QY 956 sGluLeuGluArgLysLeuSerAspLeuGlu-----LysLysMetGluThrSerHisAs 974
 QY 2842 TGAGTGTTTACCT--GAGCTACATATCAAAA-----GAAATAAGACACACAAA 2889

Db : : : : :
 QY 974 nGlnCysGlnGluLeuLysAlaArgTyrGluuArgAlaThrSerGluThrLysThrLysH 994
 QY 2890 TGGCAAAATAGAAGTCTCTCTGAAG-----CCTTCTCATTGAGCC 2934
 Db : : : : :
 QY 994 sGluGluileLeuGlnAsnLeuGlnLysThrLeuLeuAspThrGluAspLysLysG 1014
 QY 2935 TGCACCTGAATGCACAACTCTGTTCACAAATAAAGGCTTAGAATGGAAGAAATAACAAAC 2994
 Db : : : : :
 QY 1014 yAlaArgGlu--GluAsnSerGlyLeuLeuGlnGluGluLeuArgLysGlnAl 1033
 QY 2995 ATTGAGACGAGATTCACTACCTATCAAAAATCTTGATGCACTCTCTTCTGTAAG 3054
 Db : : : : :
 QY 1033 aGluLysAlaLysAlaAlaGlnThrAlaGlu-----AspAlaMetGlnIleMetGlu-- 1050
 QY 3055 AGCAAGGCAACTTAAATAAGATAACTGTGAACAAATATACAGCAAAATGGAACAAATGAA 3114
 Db : : : : :
 QY 1051 -----GlnMetThrLysGluLysThrGluThrLeu--AlaSerLeuGluAspThrL 1067
 QY 3115 AAATAAGTTTTGTGTACTACAAAGGAACTGTGAGAGCGAAGAA-----ATAAATC 3168
 Db : : : : :
 QY 1067 sGlnThrAsnAlaLysLeuGlnAsnGluLeuAspThrLeuLysGluAsnAsnLysAs 1087
 QY 3169 ACAGTTAGAGAACCAAAAGCTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTGCCTTT 3228
 Db : : : : :
 QY 1087 nValGluGluLeuAsnLysSerLys-----GluLeuLeuThrValGlu----- 1101
 QY 3229 AAATCAAGAA--GAAGAGAGAGAGAAATGTCGATATATATAA----- 3271
 Db : : : : :
 QY 1102 -AsnGlnLysMetGluGluPheArgLysGluileGluThrLeuLysGlnAlaAlaG 1121
 QY 3272 -----GAAATAATTAGACCCGAGCAAGCAACTTAG 3300
 Db : : : : :
 QY 1121 nLysSerGlnGlnLeuSerAlaLeuGlnGluGluAsnValLysLeuAlaGluLeuG 1141
 QY 3301 GAAAAAG----- 3307
 Db : : : : :
 QY 1141 yArgSerArgAspGluValThrSerHisGlnLysLeuGluGluArgSerValLeuAs 1161
 QY 3308 -----TTAGAACTGAACACACCACTTGACAGACTCTCAGATATACAGATATAGA 3357
 Db : : : : :
 QY 1161 nAsnGlnLeuLeuGluMetLysArgGluSerLysPheileLysAspAlaAspGluG 1181
 QY 3358 ATTCAAAAGTGAACAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAAGTCAAAATGA 3417
 Db : : : : :
 QY 1181 uLysAlaSerLeuGlnLysSerIleSer--IleThrSerAlaLeuLeuThrGluLysAs 1200
 QY 3418 TCTCTTTCATGAAATTTGCATGTTGAAAAGGAAATTCCTGCTTAAACTGGAAGTAGC 3477
 Db : : : : :
 QY 1200 pAlaGluLeuGluLys-----LeuArgAsnGluValThrValLeuArgGlyGluAsnAl 1218
 QY 3478 CACACTCAACATCAACACAGGTCGAGGAGAAATAATACTTTGAGGACATTAAGATT-- 3535
 Db : : : : :
 QY 1218 aSerAlaLysSerLeuHisSerValValGlnThrLeuGluSerAspLysValLysLeuG 1238
 QY 3536 -TTCAAGAAAAGAAATGCTGAACCTTCAATGACCCCTTAAACTGAAACAG 3583
 Db : : : : :
 QY 1238 uLeuLysValLysAsnLeuGluLeuGlnLysGluAsnLysArgGln 1254
 RESULT 3
 S67593
 transport protein US01 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein D2552; protein YDL058w
 C:Species: Saccharomyces cerevisiae
 C:Date: 12-Jul-1996 #sequence revision 12-Jul-1996 #text_change 21-Jul-2000
 C:Accession: S67593; A38455; S30782
 R:Bioecker, H.; Brandt, P.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67597
 A:Accession: S67593
 A:Molecule type: DNA
 A:Residues: 1-1790 <BLO>
 A:Cross-references: EMBL:Z74106; NID:g1431058; PID:e253003; PID:g1431059; MIPS:YDL058w

A; Experimental source: strain S288C
R; Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
J. Cell Biol. 113, 245-260, 1991
A; Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transp
A; Reference number: A38455; MUID: 91185402; PMID: 2010462
A; Accession: A38455
A; Molecule type: DNA
A; Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>
A; Cross-references: GB:X54378; NID: g4777; PID: CAA38253.1; PID: g4778
A; Note: the authors translated the codon ACT for residue 768 as Ile
R; Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A; Description: An integrin analogue in *Saccharomyces cerevisiae*.
A; Reference number: S30782
A; Accession: S30782
A; Molecule type: DNA
A; Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580, 'S'
A; Cross-references: EMBL: L03188
C; Genetics:
A; Gene: SGD:USO1; INT1
A; Cross-references: SGD: S0002216; MIPS: YDL058w
A; Map position: 4L
A; Keywords: coiled coil; transmembrane protein
F; 326-342/Domain: transmembrane #status predicted <TM1>
F; 394-410/Domain: transmembrane #status predicted <TM2>
F; 617-633/Domain: transmembrane #status predicted <TM3>

Alignment Scores:
Pred. No.: 1,56e-11 Length: 1790
Score: 342.00 Matches: 261
Percent Similarity: 36.30% Conservative: 180
Best Local Similarity: 21.48% Mismatches: 422
Query Match: 5.30% Indels: 352
DB: 2 Gaps: 55

US-09-602-362E-26 (1-3673) x S67593 (1-1790)

QY	584	GTAGATGTGTATGGCAACAGCGCTCCATTATGCGTTTAT-----AGT	628
Db	586	LeuAspGluThrGlyLeuProLysValTyrPheSerThrTyrPheIleGlnLeuPheAsn	705
QY	629	GAGAAATTATTAATGTGGCAACACTGCTGTCTATGGTGCAGTGCATCGAGTGGCAAAAC	688
Db	706	GluAsnIleTyrArgIleArgThrAlaLeuSerHisAspProAspGluGluProIleAsn	725
QY	689	AAGCTAGCTCACACCCCTTTTACTGGCCATACAGAAAAAGCAAGCAAACTGTGGAA	748
Db	726	LysIleSerPhe-----GluGluValGlu	733
QY	749	TTTTTACTAACAAAAATGCAATGCAACGCAATTAAATGAGTCTAAATGCAAGCCCTC	808
Db	734	LysLeu-----GlnArgGlnCysThrLysLeu	742
QY	809	ATCCTTGCCATATGTGAGGCTCATCAGAGATAGTCGGCATGCTCTTCAGCAAAATGTT	868
Db	743	-----LysGlyGluIleThrSer-----LeuGlnThrGluThr	753
QY	869	GAGCTCTTTGCTGAAGACATA-----CATGGAATA	898
Db	754	GluSerThrHisGluAsnLeuThrGluLysLeuIleAlaLeuThrAsnGluHisLysGlu	773
QY	899	ACTGCAAGCTTATGCTGCTGCTGGAGTTAATTACATTCATCAACAACTTTGGNA	958
Db	774	LeuAspGluLysTyrGlnIle-----LeuAsnSerSerHisSerSerLeuLysGlu	790
QY	959	CATATACGAAATATACCT-----AAAAATCCTCAAAATACCAATCCAGGAAGCA	1009
Db	791	AsnPheSerIleLeuGluThrGluLeuLysAsnValArgAspSerLeuAspGluMetThr	810
QY	1010	-----TCTACAGAACACCTGATGAGCTGCACCCCTTGGCGGAAGA	1051
Db	811	GlnLeuArgAspValLeuGluThrLysAspLysGluAsnGlnThrAlaLeuLeuGlu---	829

QY	1052	ACACCTGACAGCGCTGAAAGCTTGCTGCGAAAAAACACCTGACGAGCTGCACGCTTGGTG	1111
Db	830	-----TyrLysSerThrIleHisLysGlnGluAspSerIleLysThrLeuGlu	845
QY	1112	GAGGGAACGCTGCGCAAAATTTCAATGCTCTGGGAAAGCAACATCTGGA-----	1159
Db	846	LysGlyLeuGluThrIleLeuSerGlnLysLysAlaGluAspGlyIleAsnLysMet	865
QY	1160	-----AGTTTGAACAGTCAACAGAAAGAAACACCTAGAGAAA	1195
Db	866	GlyLysAspLeuPheAlaLeuSerArgGluMetGlnAlaValGluGluAsnCysLysAsn	885
QY	1196	ATTTTGAGGCTACAAAAGAAACATCTGAGAAATTTTCATGGCCAGCAAAAGAAAGATCT	1255
Db	886	Leu-----GlnLysGluLysAsp	891
QY	1256	AGGAAGATCACATGGGAGGAAAGAAACA-----TCTGTAAAGACTGAATCGTGGCAGGA	1312
Db	892	LysSerAsnValAsnHisGlnLysGluThrLysSerLeuLysGluAspIleAlaAlaLys	911
QY	1313	GTAAACACTAATAAACT-----	1330
Db	912	IleThrGluIleLysAlaIleAsnGluAsnLeuGluMetLysIleGlnCysAsnAsn	931
QY	1331	-----GAAGTTTGGAA-----AAA	1345
Db	932	LeuSerLysGluLysGluHisIleSerLysGluLeuValGluTyrLysSerArgPheGln	951
QY	1346	GGAACATCTTAATGATTGTCATGCTACAAAAGAAACATCTCAAAAGCAAGTACAAT	1405
Db	952	SerHisAspAsnLeuValAlaLysLeuThrGluLysSerLeuAlaAsnAsnTyr	971
QY	1406	GTGATGTGAGTTCTGTAGAGCCTATATTTCAGTCTTTTGGCACACGGACTATTGAAAT	1465
Db	972	LysAspMetGlnAla-----GluAsn	978
QY	1466	TCACAGTGTACAAAAGTTGAGGAAGAC-----TTTAAT	1498
Db	979	GluSerLeuIleLysAlaValGluGluSerLysAsnGluSerSerIleGlnLeuSerAsn	998
QY	1499	CTTGCTACCAAGATTATCTTAAGAGTGTGCACAGATTATACGTTTACCTGATGCT	1558
Db	999	LeuGlnAsnLysIleAspSerMetSerGlnGluLysGluAsnPheGlnIleGluArgGly	1018
QY	1559	ACATATCAAAAAGATATCAAAAACATAATCAAAAATAGAAAT-----	1603
Db	1019	SerIleGluLysAsnIleGluGlnLeuLysLysThrIleSerAspLeuGluGlnThrLys	1038
QY	1604	CAGATGTTCCCATCAGATCCAAACGAGAGAGAGATGAAGAAATATCTTGGGATCTGGG	1663
Db	1039	GluGluIleLysSerLysSerAspSerSerLysAspGluTyrGluSer-----GlnIle	1056
QY	1664	AGTCTCTTTGAGAGTTCTGCAAAAGACTCAAGTGTGTATCTACCTGAGTCTATGTATCAGAAA	1723
Db	1057	SerLeuLeuLysGluLysLeuGluThrAlaThrAlaAsnAspGluAsnValAsnLys	1076
QY	1724	GTAATGAGATAATAGAGAGTAGAGACCTTCCTGAGAGCCATCTGCTTCAGACCT	1783
Db	1077	IleSerGluLeuThrLysThrArgGluGluLeuAlaGluLeuAlaIleTyrLys---	1095
QY	1784	GCCGTGAAATGCAAAAGACTGTTCCAAATAAAGCCTTTGAAATTGAAGATGAACAAACA	1843
Db	1096	-----AsnLeuLysAsnGlu-----	1100
QY	1844	TTGAGACAGCTCAGATGTTCCCATCAGATCCAAACAAAGAGCAGCATGAGAAATCT	1903
Db	1101	---LeuGluThrLysLeuGluThrSerGluLysAlaLeuLysGluValLysGluAsnGlu	1119
QY	1904	TGGGATTTCTGAGAGTCCCTGTGAGACGGTTTCACAGAAAGGATGTGTATTATTACCCAAA	1960
Db	1120	-----GluHisLeuLysGluGluLysIleGlnLeuGluLysGlu	1132
QY	1961	GCTACACAT-----CAAAA	1975

A;Accession: S28261
 A;Molecule type: mRNA
 A;Residues: 1-2663 <YEN>
 A;Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
 C;Genetics:
 A;Gene: GDB:CENPE
 A;Cross-references: GDB:361164; OMIM:117143
 A;Map position: 4q24-4q25
 C;Superfamily: centromere protein E; kinesin motor domain homology
 C;Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 F;7-335/Domain: kinesin motor domain homology <KMT>
 F;86-93/Region: nucleotide-binding motif A (P-loop)
 F;486-2183/Domain: coiled coil #status predicted <COI>
 F;92/Binding site: ATP (Lys) #status predicted

Alignment Scores:
 Pred. No.: 7,56e-11 Length: 2663
 Score: 330.00 Matches: 250
 Percent Similarity: 37.53% Conservative: 224
 Best Local Similarity: 19.79% Mismatches: 494
 Query Match: 5.12% Indels: 295
 DB: 1 Caps: 50

US-09-602-362E-26 (1-3673) x S28261 (1-2663)

QY	478	AATGTCCTTGATGGCGAAGGAGGACACCTCTGATGAAGCTCTACATGCGAGAGGAA	537	1081	AAAAACACCTGACGAGGCTGCACGCTTGGTGGAGGACGCTCTCCCAAAATTCATATGCT	1140
DB	702	SerLeuIleAspGlyLysValProLysAspLeuLeuCysAsnLeuGluGly	721	928	rGlnGluLysAspLeuLysGlnLeuGlnGlnSerLeuGlnIleGluArgspGlnLe	948
QY	538	GTTTGTGCAATATCTCATAGATGCTGCTGATCTAAATATGATGATGATG	597	1141	GGGAAAGACACATCTGGAAAGTTTGAACAGTCA-----ACAGAAAGAACACCTAGAA	1194
DB	721	s-----IleThrAspLeuGlnLysGluLeuAsn---LysGluValGlu	735	948	uLysSerAspIleHisAspThrValAsnMetAsnIleAspThrGlnGluGlnLeuArgAs	968
QY	598	CAACAGCGCTCCATATCCCTTTATAGTAGAATTTATTAATGTCGCGCAACACTGCT	657	1195	AATTTGAGCGCTCAAAAGAAACATCTGAGAAATTT-----TCATGGCCAGC	1242
DB	735	uAsnGluAlaLeuArg-----GluGluValIleLeuLeuSerGluLeu	750	968	nAlaLeuGluSerLeuLysGlnHisGlnGluThrIleAsnThrLeuLysSerLysIleSe	988
QY	658	GTCTATGTCGTCAGTCATC-----GAGGTGCAAAACAGCTAGCTCAGCTAC	702	1243	AAAAAGAAATCTAGGAAGATCACATGGGAGGAAAAAGAAACATCTCTAAAGACTGAATG	1302
DB	750	sSerLeuProSerGluValGluArgLeuArgLysGluIleGlnAspLysSerGluGlu	770	988	rGluGluValSerArgAsnLeuHisMetGluGluAsnThrGlyGluThrLysAspGluPh	1008
QY	703	ACCCCTTTACTGCC-----ATACAGAAAGAGCAA	735	1303	C-----GTGCGAGAGTAACACCTATAAACTGAAGTTTGGAAAAAGAAACATC	1353
DB	770	uHisIleIleThrSerGluLysAspLysLeuPheSerGluValValHisLysGluSerAr	790	1008	eGlnGlnLysMetValGlyIle-----AspLysLysGlnAspLeuGluAlaLysAsnTh	1026
QY	736	G-----CAAACTGTGGAATTTTACTAACAACAAAATGC	768	1354	TAATATGATTCATGCTCCTACAAAGAAACATCTACAAAAGCAAGTACAAATGTGGATGT	1413
DB	790	gValGlnGlyLeuLeuGluGluIleGlyLysThrLysAspLeuAlaThrGlnSe	810	1026	rGlnThrLeuThrAlaAspValLysAsp-----AsnGluIleI	1039
QY	769	AATGCAACGCATTTAATGAGTCTAAATGACAGCCCTCATGCTTGCATATGTGAAG	828	1414	GAGTTCTGTAGAGCTATATTCAGTCTTTTGGCACACGG-----	1453
DB	810	rAsnTyLysSerThrAspGlnGluPheGlnAsnPheLysThrLeuHisMet-----As	828	1039	eGluGlnGlnArgLysIlePheSerLeuIleGlnGluLysAsnGluLeuGlnMetLe	1059
QY	829	CTCATCAGACATAGTCGGCATGCTTCTCAGCAAAATGTTGACGCTTTGCTCAACACAT	888	1453	-----	1453
DB	828	pPheGluGlnLysTyLysMetValLeuGlnGluAsnGluArgMetAsnGlnGluIleVa	848	1059	uGluSerValIleAlaGluLysGluGlnLeuLysThrAspLeuLysGluAsnIleGluMe	1079
QY	889	ACAT---GGAATAACTGAGACGTTATGCTGCTGCTGTGGA-----GT	930	1454	ACTATTCAAAATTCACAGTGTACAAA-----	1480
DB	848	AsnLeuSerLysGluAlaGlnLysPheAspSerSerLeuGlyAlaLeuLysThrGluLe	868	1079	tThrIleGluAsnGlnGluGluLeuArgLeuLeuGlyAspGluLeuLysLysGlnGlnG	1099
QY	931	TAATTAATCATCAACATTTTGAACATATACCAAAATACCTAAA-----	979	1481	-----GTTGAGGAAGACTTTAATCTTGTCTACCAAGATTATCTTAAGAGTCTCGACAGAA	1536
DB	868	userTyLysThrGlnGluLeuGlnLysThrArgGluValGlnGluArgLeuAsnGln	888	1099	uIleValAlaGlnGlnLysAsnHisAla-----IleLysLysGlnGluGluLeuSe	1116
QY	980	-----AATCTCAAAATACCAATCCAGAGGAACATCTACAGGAACACCTG	1026	1537	TTATACGTTTACCTGATCATATCAAAAGATATCAAAACATAATAATCACAATA	1596
DB	888	uMetGluGlnLeuLysGluGlnLeuGluAsnArgAspSerProLeuGlnThrValGluAr	908	1116	rArgThrCysAspArgLeuAlaGluValGluGluLysLysLysSerGlnGlnLe	1136
QY	1027	TGAGGTGCACCTTGGCGAAGAACACCTGCACGCGGTGAAGCTTG-----CTGA	1080	1597	AGAAGATCAGATGTTCCCATCAGAAATCCAAACGAGAGAGATCAAGAAATTTCTTGGGA	1656
DB	908	gGluLysThrLeuIleThrGluLysLeuGlnThrLeuGluGluValLysThrLeuthr	928	1136	uGlnGluLysGlnGlnGlnLeuLeuAsnValGlnGluGluMetSerGluMetGlnLysLy	1156
				1657	TTCTGGGAGTCTC-----TTTGAGAGTTCTGCAAGACTCAAGTGTGTATACCTGATC	1710
				1156	sIleAsnGluIleGluAsnLeuLysAsnGluLysLysLysGluLeuThrLeuGluHi	1176
				1711	TATGTATCAGAAAGTAATGGAG-----ATAAATAGAGAAAGTAGAGAGCTTCC	1758
				1176	sMetGluThrGluArgLeuGluLeuAlaGlnLysLeuAsnGluAsnTyArgGluGluVally	1196
				1759	TGAGAGCCATCTCCCTTCAAGCTCGCTGNGAAATCGCAAGACTGTTTCCAAAT-----	1813
				1196	sSerIleThrLysGluArgLysValLeuLysGluLeuGlnLysSerPheGluThrGluAr	1216
				1814	-----AAAGCCTTTGAATTTGAAGTAATGAACAAAC	1842
				1216	gAspHisLeuArgGlyTyIleArgGluIleGluAlaThrGlyLeuGlnThrLysGluGl	1236
				1843	ATTGAGAGCAGCTCAGATGTTCCCATCAGAAATCCAAACAAAGACGATCAAGAAATTC	1902
				1236	uLeuLysIleAlaHisIleHisLeuLysGluHisGlnGluThrIleAspGluLeuArgAr	1956
				1903	TTGGATTCCTGAGAGTCCCTGTGACGCGTTTCACAGAGCATGTGTATTATCCCAAGC	1962
				1256	gSerValSerGluLysThrAlaGlnIleIleAsnThrGlnAspLeuGluLysSerHisTh	1276
				1963	TACACATCAAAAGAAATTCGATACCTTAACTGAGAAATTTAGAGAGTCTCTCTTTAAAGA	2022

Db 1276 rYsLeuGlnGluGluProValLeuHisGluGluGlnGlu----- 1290
QY 2023 TGCTCTTGAAGCCTACCTGGAAGAAAGTTCTCTTCCAAATAAAGCCCTTGAATTT 2082
Db 1291 -----LeuLeuProAsnVal---LysLysValSerGluThrGlnGluThr-----Me 1305
QY 2083 AAAGGACAGAAACATTCAAAGCAGAGCTCTCGATAAAGATGGTCTTCTGAAGCCTAC 2142
Db 1305 tAnGluLeuGluLeuLeuThrGluGlnSerThrLysAspSer-----Th 1321
QY 2143 CTGTGAAGAAAGTTCTCTTCCAAATAAAGCCTTAGAATTTAAAG-----GACAG 2193
Db 1321 rThrLeuAlaArgLeuGluMetGluArgLeuAsnGluLysPheGlnGluSerG1 1341
QY 2194 AGAAACACTCAAAGCAGAGCTCTCTGATATGATGCTCTTGAAGCCTACCTGTGGAAG 2253
Db 1341 nGluGluLeuLysSerLeuThrLysGluArgAspAsnLeu-----LysThrIleLysG1 1359
QY 2254 GAAAGTTCTCTTCCAAATAAAGCTTTAGAAATTTGAAGCAGAGAAACATTCAAAGCAGC 2313
Db 1359 uAlaLeuGluValLysHisAspGlnLeuLysGluHisIleArgGluThrLeu-----Al 1377
QY 2314 TCAGATGTTCCCATCGAATCCAAACAA-----AAGGATGATGAAGA 2355
Db 1377 aLysIleGlnGluSerGlnSerLysGlnGlnSerLeuAsnMetLysGluLysAspAs 1397
QY 2356 AAATCTCTGGATTTTGAGAGTTCTCTGAGACTCTCTTACAGAATGATGCTGTTTACC 2415
Db 1397 nGluThrThrLysIleValSerGluMetGlnPheLysProLysAspSerAlaLeuLe 1417
QY 2416 CAAGGCTACACATCAAAAAGAAATTCGATACCTTAAGTGAATAATAGAGAGTCTCCTGA 2475
Db 1417 uArgIle-----GluIleGluMetLeuGlyLeuSerLysArgLeuGlnGluSerHisAs 1435
QY 2476 TAAGATGTCCTCTGAAGCCTACCTGTGGAATGAAATTTCTCTTCCAAATAAAGCCTT 2535
Db 1435 pGlu-----MetLysSerValAlaLysGluLys----- 1444
QY 2536 AGAATTGAAGGACAGAGAAACATTCAAAGCAGGATGTGAGTCTCTGTAGACTCCACATT 2595
Db 1445 -----AspAspLeuGlnArgLeuGlnGluValLe 1454
QY 2596 CAGCTCTTTTGGCAACCCTACTGTAATAATTCACAGTCTCAAAAGTTGAGAGACTT 2655
Db 1454 u-----GlnSerGluSerAspGlnLeuLysGluAsnI1 1465
QY 2656 TAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACCTGACACACAGAGACGTGATAT 2715
Db 1465 e-----LysGluIleValAlaLysHisLeuGluThrGluGluGluLysVa 1481
QY 2716 TGGCATATTGAACGAGCTCCACAGATCAAAATAAGATGCCCCACATCAGAATTAGG 2775
Db 1481 lAlaHisCysCysLeuLysGluGlnGluThrIleAsnGluLeuArgValAsnLeuSe 1501
QY 2776 AGAAAGAGAGATACAAATCACTTACAGATTCAGATTCTGAGATTACTCTGTGAGTGAT--AC 2832
Db 1501 rGluLysGluThrGluLysSerThrIleGlnLysGlnLeuGluAlaIleAsnAspLysLe 1521
QY 2833 ACAGAAT-----TATGAGTGTTTTACCTGAGGCTTACATATCAAAAAGAAAT 2877
Db 1521 uGlnAsnLysIleGlnGluIleTyrGlu---LysGluGlnGlnLeuAsnIleLysGlnI1 1540
QY 2878 AAAGACACAAATGGCAAAATAGACAGTCTCTGAAAAGCCTTCTCAGTTTGAGCCTGC 2937
Db 1540 eSerGluValGlnGluAsnValAsnGluLeuLysGlnPheLysGluHisArgLysAlaLys 1560
QY 2938 CACTGAAATGCCAAATCTGTTTCCAAATAAAGCCTTAGAATGGAAGATAAAA----- 2989
Db 1560 sAspSerAlaLeuGlnSerIleGluSerLysMetLeuGluLeuThrAsnArgLeuGlnG1 1580
QY 2990 -----CAAACATTGAGACAGATTCACCTACCTATCAAAAATCTTGA 3033
Db 1580 userGlnGluGluIleGlnIleMetIleLysGluLysGluGluMetLysArgValGlnG1 1600
QY 3034 TGCACTTCTCTTCTGTAAGAGAGAAAGCACTTAAAGAAAGATCACTGTGACAAATTAC 3093
Db 1600 uAlaLeu-----GlnIleGluArgAspGlnLeuLysGluAsnThrLysGluIleVa 1617
QY 3094 AGCAAAAATGGAACAAATGAAATAAAG-----TTTTGTGTACTA----- 3133
Db 1617 lAlaLysMetLysGluSerGlnGluLysGluTyrGlnPheLeuLysMetThrAlaValAs 1637
QY 3134 -----CAAAGGAATGTGATATATTAAAGAAAAAATTTAGACCGAGAGCAACTTAGGAAAAA 3186
Db 1637 nGluThrGlnGluLysMetCysGluIleGluHisLeuLysGluGlnPheGluThrGlnLys 1657
QY 3187 AGCTAAATGGGAACAAGAGCTCTGAGTGTGAGATTGCTTAAATCAAGAAGAGAGAA 3246
Db 1657 sLeuAsnLeuGlu-----AsnIleGluThrGluAs 1667
QY 3247 GAGAAGAAATGTGATATATTAAAGAAAAAATTTAGACCGAGAGCAACTTAGGAAAAA 3306
Db 1667 nIleArgLeuThrGlnIleLeuHisGluAsnLeuGluGluMetArgSerValThrLysG1 1687
QY 3307 GTTAGAAGTGAACACCAACTTGAACAGACTCTCAGAAATCAAGATATAGAAATGAAA-- 3364
Db 1687 uArgAspAspLeuArgSerValGluGluThrLeuLysValGluArgAspGlnLeuLysG1 1707
QY 3365 -----AGTCTAACAGTAATTTGAAT-----CAGGTTTCTCA 3396
Db 1707 uAsnLeuArgGluThrIleThrArgAspLeuGluLysGlnGluGluLeuLysIleValHi 1727
QY 3397 CACTCATCAAGTCAAAAT----- 3415
Db 1727 sMetHisLeuLysGluHisGlnGluThrIleAspLysLeuArgGlyIleValSerGluLys 1747
QY 3416 -----GATCTCTTTTCATGAAAATTTGCATGTTTGAAGAAA 3447
Db 1747 sThrAsnGluIleSerAsnMetGlnLysAspLeuGluHisSerAsnAspAlaLeuLysAl 1767
QY 3448 GGAAATTTGCCATGCTATAAAGTGAAGTAGCCACTGAAACATCAACACCGGTGAAGGA 3507
Db 1767 aGlnAspLeuLysIleGlnGluGluLeu---ArgIleAlaHisMetHisLeuLysGluG1 1786
QY 3508 AAATAAATACCTTTCAGGACATT----- 3529
Db 1786 nGlnGluThrIleAspLysLeuArgGlyIleValSerGluLysThrAspLysLeuSerAs 1806
QY 3530 -----AAGATTTTACAAGAAAAAATGCTGAACTTCAA----- 3562
Db 1806 nMetGlnLysAspLeuGluAsnSerAsnAlaLysLeuGlnGluLysIleGlnGluLys 1826
QY 3563 -----ATGACCTTAAACTCAACACAGAAAAACAGTAACAAAAAGGCG 3603
Db 1826 sAlaAsnGluHisGlnLeuIleThrLeuLysLysAspValAsnGluThrGlnLysLysVa 1846
QY 3604 ATCTCAGTATAGAGAGAGCTTAAAGTTCTGACCGCAGAGAACACGATGCTGACTTAA 3663
Db 1846 lSerGlu---MetGluGlnLeuLysLysGlnIleLysAspGlnSerLeuThrLeuSerLy 1865
QY 3664 ATTGAAG 3670
Db 1865 sLeuGlu 1867

RESULT 5
JC5421
smooth muscle myosin heavy chain 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001
C:Accession: JC5421
R:Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A:Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A:Reference number: JC5420; MUID:97242182; PMID:9125171
A:Accession: JC5421

A:Molecule type: mRNA
A:Residues: 1-1938 <RAS>
A:Cross-references: DBU:D85924; NID:G1945079; PIDN:BA19691.1; PID:G1945080
A:Experimental source: smooth muscle
C:Comment: This protein plays a role in smooth muscle cell contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:
Pred. No.: 1,47e-10 Length: 1938
Score: 324.50 Matches: 227
Percent Similarity: 37.07% Conservative: 190
Best Local Similarity: 20.18% Mismatches: 421
Query Match: 5.03% Indels: 288
DB: 2 Gaps: 40

US-09-602-362E-26 (1-3673) x JCS421 (1-1938)

Qy	951	TTTTGGAAATATACGAAATTTACCTAAATAATCTCAAAATACCAATCCAGAGGAACAT	1010	Db	1092	LeuGlnAlaLeuAlaArgLeuAspGluGluLeuAlaGlnLysAsnAsnAlaLeu---	1110
Db	834	TrpTrpArgLeuPheThrLysValLysProLeuLeuGlnValThrArgGlnGluGlu	853	Qy	1553	GATGCTACATATCAAAAAGATATCAAAACAATAAATCAAAAATAGAGATCAGATGTT	1612
Qy	1011	CTACAG-----GAACACCTGATGAGCTGCACCTTGGCGGAAACACCTGACA	1061	Db	1111	-----LysLysIleArgGluLeuGluGlyHisLeuSerAspLeuGlnGlu	1125
Db	854	MetGlnAlaLysGluGluGluMetGlnLysIleThrGluArgGlnGlnLysAlaGluThr	873	Qy	1613	CCATCAGAATCCAAA-----CGAGAGGAGATCAAGAATATTCTTCGGATTCTGGG	1663
Qy	1062	CGCTGAAACCTTCTCGGAAAAACACCTG-----	1091	Db	1126	AspLeuAspSerGluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGly	1145
Db	874	GluLeuLysGluLeuGluGlnLysHisThrGlnLeuAlaGluGluLysThrLeuLeuGln	893	Qy	1664	AGTCTCTTTGAGAGTTCTGCAAGACCTCAAGTG-----TGTATACCTGAGTCTATGAT	1717
Qy	1092	-----ACGAGCTGCAGCTTGTGGAGGAAAGCTCTGCCAAATTC	1133	Db	1146	GluGluLeuGlu-----AlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThr	1164
Db	894	GluGlnLeuGlnAlaGluThrGluLeuTyrAlaGluSerGluGluMetArgValArgLeu	913	Qy	1718	CAGAAAGTAATGGAGATAAATAGAGAA-----	1744
Qy	1134	ANTGCTGGGAAAGACACATCTGGAA-----	1160	Db	1165	GlnGlnGluLeuArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAsp	1184
Db	914	AlaAla-----LysLysGlnGluLeuGluGluLeuLeuHisGluMetGluAlaArgLeuGlu	932	Qy	1744	-----	1744
Qy	1161	-----AGTTTCAACAGCTCAACAGAGAAACA-CCTAGGAAAT	1198	Db	1185	GluGluThrArgSerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAla	1204
Db	933	GluGluCluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMet	952	Qy	1745	GTAGAAGAGCTTCTGTGAGAGCCATCTGCTTCAAGCCTCCGTN-----GAA	1792
Qy	1199	TTGAGGCTCAAAAGAAACATCTGAGAAATTTTCATGCGCCACCAAAAGAGATCTAGG	1258	Db	1205	ValGluGluLeuThrGluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLys	1224
Db	953	LeuAspLeuGluGlnGlnLeuGluGluGluAlaAlaArgGlnLysLeuGlnLeuGlu	972	Qy	1793	ATGCAAAAGACTGTTTCCAAAATAAAAGCCTTTCAATTGAAGATGAACAAACATTCAGAGCA	1852
Qy	1259	AAGATCATGGGAG-----	1273	Db	1225	SerLysGlnThrLeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGly	1244
Db	973	LysValThrAlaGluAlaLysIleLysLysLeuGluAspIleLeuValMetAspAsp	992	Qy	1853	GCTCAGATGTTCCCATCAGATCCAAACAAAGACCATCAAGAAATTTCTTGGGATCT	1912
Qy	1274	-----GAAAGAAACATCTGTAAGACTGAATGCTGTCGAGAGTAACA	1318	Db	1245	GlnAlaLysGlnGluValGluHisLysLysLysLeuGluValGlnLeuGlnAspLeu	1264
Db	993	GlnAsnSerLysLeuSerLysGluArgLysLeuGluGluArgValSerAspLeuThr	1012	Qy	1913	GAGAGTCCCTGTGAGACGGTTTACAGAGAGTGTGTATTACCCAAAGCTACACAT---	1969
Qy	1319	CCTAATAAAGCTAAAGTTTGGAAAAAGGACATCTAATATGATTCATCTCAAAA	1378	Db	1265	GlnSerLysCysSerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLys	1284
Db	1013	ThrAsnLeuAlaGluGluGluGluLys-----AlaLysAsnLeuThrLysLysSerLys	1031	Qy	1970	---CAAAAGAAATTCGATACCTTAAGTGGAAATATAGAGAGTCTCTCTGTAAGATGGT	2026
Qy	1379	GAAACATCTCAAAAGCAAGTACAAATGTG-----	1408	Db	1285	LeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAla-----	1299
Db	1032	HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluLysSerArgGln	1051	Qy	2027	CTTCTGAGCCTACTCTGTGGAGGAAAGTTTCTCTCCAAATAAGCCTTAGAATTA---	2083
Qy	1409	-----GATGTAGTCTCTGA-----	1432	Db	1300	-----GluGlyLysAlaIleLysLeuAla	1307
Db	1052	GluLeuGluLysLysArgLysLeuGluGlyAspAlaSerAspPheHisGluGlnIle	1071	Qy	2084	AAGGACAGAGAAACATTCAAAGCAGAGTCTCTGTAAGATGGTCTCTGAAAGCCTACC	2143
Qy	1433	TTCAAGTCTTTTGGCACACGAGCTATTGAAATTTACAGATGTACAAAGTGTGAGGAGAC	1492	Db	1308	LysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGlnGluLeuGlnGluGlu	1327
Db	1072	AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluGlu	1091	Qy	2144	TGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATC	2203
Qy	1493	TTTAATCTCTACCAAGATTATCTCTAGAGTGTGCAGAGATTATACGTTTACCT	1552	Db	1328	ThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSer	1347

Db 1403 GlyLeuSerGlnGlnThrGluGluysAlaAlaIaTyAspLysLeuGluLysThrLys 1422
 QY 2480 GATGGTCTTCTGAGCCTACCTGCGAATGAAATTTCTCTCCAAATAAGGCC----- 2533
 Db 1423 AsnArgLeuGlnGlnGluLeuAspAspLeuValAspLeuAspAsnGlnArgGlnLeu 1442
 QY 2534 -----TTAGATTGAAGACAGA----- 2551
 Db 1443 ValSerAsnLeuGluLysLysGlnLysPheAspGlnLeuLeuAlaGluLysAsn 1462
 QY 2552 -----GAAACATTCAAAGCAGAGAGTGTGAGTCTCTAGAGTCC 2590
 Db 1463 IleSerSerLysTyAlaAspGlnArgAspAlaGluAlaGluArgGlnLysGlu 1482
 QY 2591 ACATTCAAGTCTTTTGGCAACCGACTACTGAAAATTCACAGTCTACAAAAGTTCAGGAA 2650
 Db 1483 ThrLysAlaLeuSerLeuAlaArgAlaLeuGlu-----GluAlaLeuGluAlaLysGlu 1500
 QY 2651 GACTTTAATCTTACTACCAAGAG-----GGAGCAACAAG 2686
 Db 1501 GluLeuGluArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLys 1520
 QY 2687 ACAGTAACCTGGACAACAGGAACTGATATTGGCATTATTGAACGAGCTCCACAAGATCAA 2746
 Db 1521 AspAspValGlyLysAsnValHisGluLeuGluLysSerLysArgAlaLeuGluThrGln 1540
 QY 2747 ACAATAAGATCCCATCGAATTAGGAAGAAAGAGATACAAAATCAATCTTCAGAT 2806
 Db 1541 MetGluGluMetLysThrGlnLeuGluGluSerGluAspValGlnAlaThrGluAsp 1560
 QY 2807 TCTGAGATT--ATCTCTGTGATGATACACAGATTATTGAGTGTTCCTGAGGCTACA 2863
 Db 1561 AlaLysLeuArgLeuGluVal-----AsnMetGlnAlaLeu-----LysGlyGln 1575
 QY 2864 TATCAAAAAGAAATAAGACACAAATGGCAAAATAGAGATGCTCCT-----GAA 2914
 Db 1576 PheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluLysArgArgGlnLeuGln 1595
 QY 2915 AAGCTTCTCATTGAGCTGCCACTGAAATGCAAACTGTCTCCAAATAAAGCTTA 2974
 Db 1596 ArgGlnLeuHis---GluTyGlnThrGluLeuGluAspGluArgLysGlnArgAlaLeu 1614
 QY 2975 GAATGGAAGAATAAACAACATTCAGAGCAGATTCACCTATCAAAATCTTGAT 3034
 Db 1615 AlaAlaAlaAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1634
 QY 3035 GCATCTCTCTTGTGGAAGAGAGAGGAACTTAAAGAAATACTGTGAAACAATTCAC 3094
 Db 1635 -----SerAlaIleLysGlyArgGlu-----GluAlaIleLysGlnLeuArg 1648
 QY 3095 GCAAAAATCGAACAATAAATAAGTTTGTGTACTACAAAAGAACTGTCAAGCG 3154
 Db 1649 LysLeuGlnAlaGlnMetLysAsp-----PheGlnArgGluLeuAspAspAla 1664
 QY 3155 AAAGAAATAAATCACTAGTTA-----GAGAACCAAAAA-----GCTAAA 3193
 Db 1665 ArgAlaSerArgAspGluIlePheAlaThrSerLysGluAsnGluLysLysAlaLysSer 1684
 QY 3194 TGGGAACAAGAGCTCTGAGTGTGAGATTGCTTTAATCAAGACAGAGAGAGAGA 3253
 Db 1685 LeuGluAlaAspLeuMetGlnLeuGlnGluAspLeuAlaAlaGluArgAlaArgLys 1704
 QY 3254 AATGTCGATATATTAAGAAAGAAAATTAGACCCGAA----- 3289
 Db 1705 GlnAlaAspLeuGluLysGluGluLeuAlaGluLeuAlaSerSerLeuSerGlyArg 1724
 QY 3290 -----GAGCAACTTAGGAAAAGTTAGAGTGAAA---CACCACTTGAACAGACT 3337
 Db 1725 AsnThrLeuGlnAspGluLysArgLeuGluAlaArgIleAlaGlnLeuGluGlu 1744
 QY 3338 CTCAGAATAACAATATAGATTGAAAAGTGTA-----ACAAGT 3376
 Db 1745 LeuGluGluGlnGlnGlnMetGluAlaMetSerAspValArgLysAlaThrLeu 1764

QY 3377 AATTGGAATCAGGTTTCTCAGCTCATGAAAGTGAAATGATCTCTTTTCATGAAAT--- 3433
 Db 1765 GlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArgSerThrAlaGlnLysAsnGlu 1784
 QY 3433 ----- 3433
 Db 1785 SerAlaArgGlnGlnLeuGluArgGlnAsnLysGluLeuArgSerLysLeuGlnVal 1804
 QY 3434 -----TGCATGTTGAAAAGGAAATTCCTCATGCTTAAACCTGGAAGTAGCC 3478
 Db 1805 GluGlyAlaValLysAlaLysLeuLysSerThrValAlaAlaLeuGluAlaLysIleAla 1824
 QY 3479 ACATGAAACATCAACACAGGTCAGGAAATAATACTTTGAGGACATTAAGATTTTA 3538
 Db 1825 GlnLeuGluGlnValGlnGlnGluAlaArgGluLysGlnAlaAlaThrLysSerLeu 1844
 QY 3539 CAGAAAAGATGCTGAACCTTCAATGACCTTAACTGAAACAGAAAACAGTAACAAAA 3598
 Db 1845 LysGlnLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1863
 QY 3599 AGGCGATCTCAGTATAGAGACAGCTTAAAGTTCTGACCGCAGAGAACACGATGCTGACT 3658
 Db 1864 MetAlaGluGlnTyLysGluAlaGluLysGlyAsnThrLysValLysGlnLeuLys 1883
 QY 3659 TCTAAATGAAGNA 3673
 Db 1884 ArgGlnLeuGluGlu 1888
 RESULT 6
 JC5420
 smooth muscle myosin heavy chain 1 - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 02-Feb-2001
 C/Accession: JC5420
 R/Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
 Biochem. Biophys. Res. Commun. 232, 313-316, 1997
 A/Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
 A/Reference number: JC5420; MUID:97242182; PMID:9125171
 A/Accession: JC5420
 A/Molecule type: mRNA
 A/Residues: 1-1972 <HAS>
 A/Cross-references: DDBJ:D85923; NID:g1945077; PIDN:BAA19690.1; PID:g1945078
 A/Experimental source: smooth muscle
 C/Comment: This protein plays a role in smooth muscle cell contraction.
 C/Superfamily: myosin heavy chain; myosin motor domain homology
 C/Keywords: nucleotide binding; P-loop
 F:88-771/Domain: myosin motor domain homology <WMOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 Alignment Scores:
 Pred. No.: 1,48e-10 Length: 1972
 Score: 324.50 Matches: 227
 Percent Similarity: 37.07% Conservative: 190
 Best Local Similarity: 20.18% Mismatches: 421
 Query Match: 5.03% Indels: 288
 DB: 2 Gaps: 40
 US-09-602-362E-26 (1-3673) x JC5420 (1-1972)
 QY 951 TTTTGGACATATACGAAATATACCTAAATCTCTAAATATCAATCCAGAGGAACAT 1010
 Db 834 TrpTrpArgLeuPheThrLysValLysProLeuLeuGlnValThrArgGlnGluGlu 853
 QY 1011 CTACAG-----GAACACCTGATGAGCTGCACCTTTGGCGGAAAGAACACCTGACA 1061
 Db 854 MetGlnAlaLysGluGluGluMetGlnLysIleThrGluArgGlnGlnLysAlaGluThr 873
 QY 1062 CGGCTGAAAGCTTCTCGAAAACACCTG----- 1091
 Db 874 GluLeuLysGluLeuGluGlnLysHisThrGlnLeuAlaGluGluLysThrLeuLeuGln 893
 QY 1092 -----ACGAGGCTCGACGCTTGTGGAGGGAACGCTCTGCCAAAATTC 1133

Db 894 GluGlnLeuGlnAlaGluThrGluLeuTyrAlaGluSerGluGluMetArgValArgLeu 913
QY 1134 AATGCTCTGGGAAAGCAACATCTGGAA- - - - - 1160
Db 914 AlaAla- - - - - LysLysGlnGluLeuGluGluIleLeuHisGluMetGluAlaArgLeuGlu 932
QY 1161 - - - - - AGTTTGAACAGTCAACAGAGAAACA-CTAGGAAAATT 1198
Db 933 GluGluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMet 952
QY 1199 TTGAGGCTCAAAAGAAACATCTGAGAAATTTCTGCGCCAGCAAAAGAAATCTAGG 1258
Db 953 LeuAspLeuGluGlnGluGluGluGluAlaAlaArgGlnLysLeuGlnLeuGlu 972
QY 1259 AGATCACTGGGAG- - - - - 1273
Db 973 LysValThrAlaGluAlaLysIleLysLysLeuGluAspLysLeuValMetAspAsp 992
QY 1274 - - - - - GAAAGAAACATCTGTAAGACTGAATGCGTGGCAGGTAACA 1318
Db 993 GluAsnSerLysLeuSerLysGluArgLysLeuGluGluArgValSerAspLeuThr 1012
QY 1319 CCTAATAAACTGAAGTTTGGAAAAAGGACATCTAATATGATTCATGCTCTACAAA 1378
Db 1013 ThrAsnLeuAlaGluGluGluLys- - - - - AlaLysAsnLeuThrLysLysSerLys 1031
QY 1379 GAAACATCTCAAAAGCAAGTCAAAATGTG- - - - - 1408
Db 1032 HisGluSerMetIleSerGluLeuGluValArgLeuLysLysGluLysSerArgGln 1051
QY 1409 - - - - - GATGCTGAGTTCTGTA- - - - - GAGCCTATA 1432
Db 1052 GluLeuGluLysLysLysLysLysLysGluGluLysAlaSerAspPheHisGluGlnIle 1071
QY 1433 TTCAGTCTTTTGGCACCGGACTATTGAAATTCACAGTGTACAAAAGTTGAGGAGAC 1492
Db 1072 AlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGlu 1091
QY 1493 TTTAATCTTCTACCAAGATTATCTCTAAGAGTGTGACAGAAATATACGTGTTACCT 1552
Db 1092 LeuGlnAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAlaLeu- - 1110
QY 1553 GATGCTACATACAAAAGATATCAAAACAATAAATCAAAAATAGAGATCAGATGTC 1612
Db 1111 - - - - - LysLysIleArgGluLeuGluGluLysIleSerAspLeuGlnGlu 1125
QY 1613 CCATCAGATCCAAA- - - - - CGAGAGGAGATGAAGATATCTTGGGATCTGGG 1663
Db 1126 AspLeuAspSerGlnArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGly 1145
QY 1664 AGTCTCTTTCAGAGTCTTCCAAAGACTCAAGTG- - - - - TGTATACCTGAGTCTATGAT 1717
Db 1146 GluGluLeuGlu- - - - - AlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThr 1164
QY 1718 CAGAAAGTAATGAGATAATACAGAA- - - - - 1744
Db 1165 GlnGlnLeuArgAlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAsp 1184
QY 1744 - - - - - 1744
Db 1185 GluGluThrArgSerHisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAla 1204
QY 1745 GTAGAGAGCTTCTGAGAGCCATCTGCTTCAAGCCTGCCGTN- - - - - GNA 1792
Db 1205 ValGluGluLeuThrGluGlnGlnGluGlnPheLysArgAlaLysAlaAsnLeuAspLys 1224
QY 1793 ATCAAAAGACTGTTCCAAATAAGCTTTGAATTCGAAGAAATGAACAAACATGAGAGCA 1852
Db 1225 SerLysGlnThrLeuGluLysGluAsnAlaAspLeuAlaGlyGluLeuArgValLeuGly 1244
QY 1853 GCTCAGATGTTCCATCAGAAATCCAAACAAAGGACGATGAAGAAATTTCTTGGGATCT 1912

Db 1245 GlnAlaLysGlnGluValGluHisLysLysLysLysLysLeuGluValGlnLeuGlnAspLeu 1264
QY 1913 GAGAGTCCCTGTGAGAGCGTTTTCACAGAGATGTGTATTATCCCAAGACTACACAT- - - 1969
Db 1265 GlnSerLysCysSerAspGlyGluArgAlaArgAlaGluLeuSerAspLysValHisLys 1284
QY 1970 - - - - - CAAAAGAAATTCGATACCTTAAGTGGAAATTAAGAGAGTCTCTCTTAAAGATGCT 2026
Db 1285 LeuGlnAsnGluValGluSerValThrGlyMetLeuAsnGluAla 1299
QY 2027 CTTCTGAAGCCTACCTCTGGAGAAAGTTTCTTCCAAATAAAGCCTTAGAATTA- - - 2083
Db 1300 - - - - - GluGlyLysAlaIleLysLeuAla 1307
QY 2084 AAGCAGACAGAAACATTCAAAGCAGAGTCTCCTGATAAAGATGCTCTTCTGAAGCCTACC 2143
Db 1308 LysAspValAlaSerLeuGlySerGlnLeuGlnAspThrGlnLeuLeuLeuGlnGluGlu 1327
QY 2144 TGTGGAAGGAAGTTTCTTCCAAATAAAGCCCTTAGAATTAAGGACAGAGAAACATC 2203
Db 1328 ThrArgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluA:GAsnSer 1347
QY 2204 AAACAGAGTCTCTGATATGATGTCTTCTGAAGCCTACCTGTGGAAGGAAGTTTCT 2263
Db 1348 LeuGlnAspGlnLeuAspGluMetGluAlaLysGlnAsnLeuGluArgHisValSer 1367
QY 2264 CTTCCAAATAAAGCTTTAGAAATTGAAGCAGAGAAACATTCAAAGCAGCTCAGATGTT 2323
Db 1368 ThrLeuAsnIleGlnLeu- - - - - 1373
QY 2324 CCATCAGATCCAAACAAAGAGATGATGAAGAAATTCCTGGGATTTTGAGATTTCCCT 2383
Db 1374 - - - - - SerAspSerLysLysLys- - - - - LeuGlnAspPheAlaSerThrIle 1387
QY 2384 GAGACTCTTTACAGAAATGATGTGTGTACCAAGCTACACATCAAAAGAAATTCGAT 2443
Db 1388 GluValMetGluGluGly- - - - - LysLysArgLeuGlnLysGluMetGlu 1402
QY 2444 ACCTTAAGTGGAAATTAAGAGAG- - - - - TCTCTGATATA 2479
Db 1403 GlyLeuSerGlnGlnTyrGluGluLysAlaAlaLysAspLysLeuGluLysThrLys 1422
QY 2480 GATGCTCTTGAAGCCTACCTGTGGAATGAAATTTCTTCCAAATAAAGC- - - - 2533
Db 1423 AsnArgLeuGlnGlnLeuAspLeuValValAspLeuAspAsnGlnArgGlnLeu 1442
QY 2534 - - - - - TTAGAATTCAGAGCAGAG- - - - - 2551
Db 1443 ValSerAsnLeuGluLysLysGlnLysPheAspGlnLeuLeuAlaGluGluLysAsn 1462
QY 2552 - - - - - GAAACATTCAGAGCAGAGATGTGATTCAGAGCTCCACAGATCAA 2590
Db 1463 IleSerSerLysTyrAlaAspGluArgAspAlaGluAlaGluAlaArgGluLysGlu 1482
QY 2591 ACATTCAGTCTTTTGGCAAAACCCACTACTCAAAATTCACAGTCTACAAAGTTGAGGAA 2650
Db 1483 ThrLysAlaLeuSerLeuAlaArgAlaLeuGlu- - - - - GluAlaLeuAlaLysGlu 1500
QY 2651 GACTTTAATCTTACTACAGGAG- - - - - GGAGCAACAAAG 2686
Db 1501 GluLeuGluArgThrAsnLysMetLeuLysAlaGluMetGluAspLeuValSerSerLys 1520
QY 2687 ACAGTAACGTGACAAACAGGAACGTGATTTGGCATTTATTGAACGAGCTCCACAGATCAA 2746
Db 1521 AspAspValGlyLysAsnValHisGluLeuLysSerLysArgAlaLeuGluThrGln 1540
QY 2747 ACAATAAGATGCCCAATCAGATTCAGGAAGAAAGAAAGATACAAATCAACTTCAGAT 2806
Db 1541 MetGluGluMetLysThrGlnLeuGluGluSerGluAspValGlnAlaThrGluAsp 1560
QY 2807 TCTCAGATT- - - - - ATCTCTGTGAGTGATACACAGAAATTTATGAGTGTTCCTCAGGCTTACA 2863
Db 1561 AlaLysLeuArgLeuGluVal- - - - - AsnMetGlnAlaLeu- - - - - LysGlyGln 1575


```
QY 2864 TATCAAAAGAAATAAAGACACAAATGGCAAAATAGAGAGTCTCCT-----GAA 2914
Db 1576 PheGluArgAspLeuGlnAlaArgAspGluGlnAsnGluGluLysArgGlnLeuGln 1595
QY 2915 AAGCTTCTCAGTTCAGCTGCGCTGCACTGAAATGCAAAATCTGTTCCAAATAAAGCTTTA 2974
Db 1596 ArgGlnLeuHis---GluThrGluThrGluGluAspGluArgLysGlnArgAlaLeu 1614
QY 2975 GAATGGAAGAAATAAACAACATTGAGAGCAGATTCAACTACCTATCAAAATCTTGGAT 3034
Db 1615 AlaAlaAlaLysLysLysLeuGluGlyAspLeuLysAspLeuLysGlnAlaAsp 1634
QY 3035 GCATCTCTCTTGTGCAAGAGGAGGAACTTTAAAAAGATAACTGTGAACAAATTTACA 3094
Db 1635 -----SerAlaIleLysGlyArgGlu-----GluAlaIleLysGlnLeuArg 1648
QY 3095 GCAAAATGCAAAATGAAATAAGTTTGTGTACTACAAAGAAAGAACTGTCAGAGCG 3154
Db 1649 LysLeuGlnAlaGlnMetLysAsp-----PheGlnArgGluLeuAspAla 1664
QY 3155 AAAGAAATAAAATCACAGTTA-----GAGAACCAAAAA-----GCTAAA 3193
Db 1665 ArgAlaSerArgAspGluIlePheAlaThrSerLysGluAsnGluLysLysAlaLysSer 1684
QY 3194 TGGGAACAAGAGCTGCGAGTGTGAGATTGCCCTTTAAATCAAGACAGAGAGAGAGA 3253
Db 1685 LeuGluAlaAspLeuMetGlnLeuGlnGluAspLeuAlaAlaAlaGluArgAlaArgLys 1704
QY 3254 AATGTCGATATATTAAGAAAGAAAAATTAGACCCGAA----- 3289
Db 1705 GlnAlaAspLeuGluLysGluGluLeuAlaGluLeuAlaSerSerLeuSerGlyArg 1724
QY 3290 -----GAGCAACTTAGGAAAAAGTTAGAGTGAAA---CACCAACTGACAGACT 3337
Db 1725 AsnThrLeuGlnAspGluLysArgArgLeuGluAlaArgIleAlaGlnLeuGluGlu 1744
QY 3338 CTCAGATAACAAGATATAGAAAGTGAAGTGTA-----ACAAGT 3376
Db 1745 LeuGluGluGluGlnGlyAsnMetGluAlaMetSerAspArgValArgLysAlaThrLeu 1764
QY 3377 AATTGAAATCAGGTTTCTCACACTCATGAAAGTGAAGATGATCTCTTTTCATGAAAT 3433
Db 1765 GlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArgSerThrAlaGlnLysAsnGlu 1784
QY 3433 ----- 3433
Db 1785 SerAlaArgGlnGlnLeuGluArgGlnAsnLysGluLeuArgSerLysLeuGlnGluVal 1804
QY 3434 -----TGCATGTTGAAAAGGAAATGCCATGCTAAACTGGAAGTAGCC 3478
Db 1805 GluGlyAlaValLysAlaLysLysSerThrValAlaAlaLeuGluAlaLysIleAla 1824
QY 3479 ACATGAAACATCAACACAGGTGAAGGAAATAAATACTTTGAGGACATTAAGATTTTA 3538
Db 1825 GlnLeuGluGluGlnValGluGlnGluAlaArgGluLysGlnAlaAlaThrLysSerLeu 1844
QY 3539 CAAGAAAGAATGCTGAATCTCAATGACCTTAAACTGAAACCTGAAACAGAAACAGTACAAA 3598
Db 1845 LysGlnLysAspLysLysLysLysGluValLeu---LeuGlnValGluAspGluArgLys 1863
QY 3599 AGGGCATCTCAGTATAGAGCAGGTTTAAAGTTCTGACGGCAGAGAACACGATGCTGACT 3658
Db 1864 MetAlaGluGlnTyrlsGluGlnAlaGluLysGlyAsnThrLysValLysGlnLeuLys 1883
QY 3659 TCTAAATTGAAGAA 3673
Db 1884 ArgGlnLeuGluGlu 1888
```

RESULT 7

T14156

kinesin-related protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T14156
R;Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
Cell 91, 357-366, 1997
A;Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromoso
A;Reference number: Z17893; MUID:98028574; PMID:9363944
C;Accession: T14156
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2954 <WOO>
A;Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1
C;Genetics:
A;Gene: XENP-E
C;Superfamily: centromere protein E; kinesin motor domain homology

Alignment Scores: 2.93e-10 Length: 2954
Score: 319.50 Matches: 251
Percent Similarity: 37.32% Conservative: 211
Best Local Similarity: 20.27% Mismatches: 397
Query Match: 4.96% Indels: 380
DB: 2 Gaps: 54

US-09-602-362E-26 (1-3673) x T14156 (1-2954)

```
QY 759 CAAAAATGCAAAATGCAAAATTAATGAGTCTAAATGCACAG----- 803
Db 1168 GlnValIleAspLeuAsnThrGlnLeuGlnSerLeuGlnAlaGlnSerIleGluLysSer 1187
QY 804 -----CCCTCATGC-----TTG 815
Db 1188 AspLeuGlnLysProLysGlnAspLeuGluGluGlyValLysLeuLeuGluMet 1207
QY 816 CCATATGTGAGGCTCATCAGAGATAGTCGSCATGCTTTCACAAAAATG----- 866
Db 1208 GluLeuLysGlyHisLeuThrAspSerGlnLeuSerIleGluLysLeuGlnGlu 1227
QY 867 -----TTGACGCTCTTTCGACACACATACATGAATACTGACGAGACCTTATGCTG 917
Db 1228 AsnLeuGluValThrGluLysLeuGlnThrLeuGlnGluMetLysAsnIleThrIle 1247
QY 918 CTGCTCGTGAAGTTAATTAATTCATCAACAACCTTTTGGAAACATATACGAAAAATACCTA 977
Db 1248 GluArgAsnGluLeuGlnThr-----AsnPheGluAspLeuLysAlaGluHisAsp 1264
QY 978 AAAATCTCAAAATPACCAATCCAGAGAAACATCTACAGGAACACCTGATGAGCTGCAC 1037
Db 1265 SerLeu-----LysGlnAspLeuSerGluAsnIleGluGlnSerIle 1278
QY 1038 CTTGGCGGAAAGAACACCTGACACGGCTGAAGCTTCTGTGAAA-----AACACCTG 1091
Db 1279 GluThrGlnAspGlu-----LeuArgAlaAlaGlnGlnGluLeuArgGluGln 1294
QY 1092 ACGAGGCTGCACGCTTGTGTGAGGGAACGCTGCCAAAAATTCATGTCTG---GGAAG 1148
Db 1295 LysGlnLeuValAspSerPheArgGlnGlnLeuAspCysSerValGlyIleSerSer 1314
QY 1149 CAACATCTGGAAGTTGAACAGTCAACAGAGAAACACCTAGGAAAATTTTGGGCGCTA 1208
Db 1315 ProAsnHisAspAlaValAlaAsnGlnGluLysValSerLeuGlyGluValAsnSerLeu 1334
QY 1209 CAA----- 1211
Db 1335 GlnSerGluMetLeuArgGlyGluArgAspGluLeuGlnThrSerCysLysAlaLeuVal 1354
QY 1212 -----AAGAAACATCTGAGAAAATTTTCATGCCCGCAAGAAA-GAAGA 1252
Db 1355 SerGluLeuGluLeuArgAlaHisValLysSerValGluGlyGluAsnLeuGluIle 1374
QY 1253 TCTAGGAAGATCACATGGGAGGAAAAGAACACTCTCTTAAGACTGAATCGGTGGCAGGA 1312
Db 1375 ThrLysLysLeuAsnGlyLeuGluLysGluIleLeuGlyLysSerGlu----- 1390
```


QY 3494 -----CACAGGTGAAGAAATAAATCTTTGAG---GAC 3526
Db 1996 AspAlaGluGlnLysAlaLysAsnGlnInLeLysGluAsnLysLysIleGluThrGlu 2015
QY 3527 ATTAAGATTTCACAGCAAGAAATGCTGAATCTCAATACACCTAAACCTGAAACACAGAAA 3586
Db 2016 LeuLysGluTyArgLysLysPheGlyGluSerGluLysThrLysThrLysGluPheLeu 2035
QY 3587 ACAGTAACAAAAGGCGATCTCAGTATAGA-----GAGCAG 3622
Db 2036 ValValGluLysLeuGluThrAspTyLysAArgAlaLysLysGluAlaAlaAspGluGln 2055
QY 3623 CTTAAAGTTCTGACGCGCAGACACAGATG 3652
Db 2056 GlnGlnArgLeuThrValGluAsnAspLeu 2065

RESULT 9
A41604
myosin heavy chain, smooth muscle, long splice form - rabbit
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A41604; PMID:92073350; PMID:1961735
R:BabiJ, P.; Kelly, C.; Periasamy, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
A:Title: Characterization of a mammalian smooth muscle myosin heavy-chain gene: complete
A:Reference number: A41604; PMID:92073350; PMID:1961735
A:Accession: A41604
A:Molecule type: mRNA
A:Residues: 1-1972 <BAB>
A:Cross-references: GB:M77812
R:Nagai, R.; Kuro-O, M.; BabiJ, P.; Periasamy, M.
J. Biol. Chem. 264, 9734-9737, 1989
A:Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN
A:Reference number: A33501; PMID:89255535; PMID:2722872
A:Accession: A33501
A:Molecule type: mRNA
A:Residues: 1455-1972 <NAG>
A:Cross-references: GB:J04833; NID:G168519; PIDN:AAA31407.1; PID:G165520
A:Experimental source: smooth muscle
A>Note: examination by Southern blotting for the regions of difference between this isofo
active splicing
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:559-572/Region: actin binding #status predicted
F:633-647/Region: actin binding #status predicted
F:844-1938/Domain: coiled coil #status predicted <COI>
F:844-1284/Region: S2
F:1285-1972/Region: light meromyosin
F:1939-1972/Domain: carboxyl-terminal <CBT>
F:129/Modified site: N6, N6, N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:701,711/Active site: Cys #status predicted

Alignment Scores:
Pred. No.: 4,97e-10 Length: 1972
Score: 315.00 Matches: 227
Percent Similarity: 37.99% Conservative: 203
Best Local Similarity: 20.05% Mismatches: 460
Query Match: 4.89% Indels: 242
DB: 1 Gaps: 41

US-09-602-362E-26 (1-3673) x A41604 (1-1972)

QY 677 GAGGTCAAAACAGCGCTCAGCCCTTTTACTGGCCATACAGAAAGACGACG 736
Db 866 GluArgGlnGlnLysAlaGluSerGlu-----LeuGlnGluLeuGlnGlnLysHisThr 883
QY 737 CAAACTGCGAA-----TTTACTACAAAATAATGCAATGCAACGCAATTAATGAG 790

Db 884 GlnLeuSerGluGluLysAsnLeuLeuGlnGlnLeuGlnAlaGluThrGluLeuTy 903
QY 791 TCTAAATGCACAGCCCTCANGCTTGGCCATATGTGAAGGCTCATCAGAGATAGTCGGCATG 850
Db 904 AlaGluAlaGluGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGluLeu 923
QY 851 CTTCTTCCAGCAAAATGTTGACGCTCTTTCCTGCAAGACATACATGCA-----ATACTGCA 904
Db 924 LeuHisGluMetGluAlaArgLeuGluGluGluAspArgGlnGlnGlnAla 943
QY 905 GAACGTTATGCTGCTGCTGCTGAGTTAATATACATTTCATCAACAACTTTTGGAAACAT--- 961
Db 944 GluArgLysLysMetAlaGlnMetLeuAspLeuGluGlnLeuGluGluGluGlu 963
QY 962 -----ATACGAAAATTA 973
Db 964 AlaAlaArgGlnLysLeuGlnLeuLysValThrAlaGluAlaLysIleLysLeu 983
QY 974 CCT-----AAAAATCCTCAAAATACCAATCCAGAAAGAACATCTACAGGA 1018
Db 984 GluAspAspIleLeuValMetAspAspGlnAsnLys----- 996
QY 1019 ACACCTGATGAGGCTGCACCCCTTGGCGAAAGAACACCTGACCGCTGAAGCTTCTGTG 1078
Db 997 LeuSerLysGluArgLysLeuLeuGluGluArgIleSerAspLeuThrThrAsnLeuAla 1016
QY 1079 GAAAAAACACCTGACGAGGCTGCACGCTGCTGGAGGGAACGCTGCGCAAAATTCATGT 1138
Db 1017 GluGlu---GluGluLysAlaLysAsnLeuThrLys----- 1027
QY 1139 CTGGGGAAGCAACATCTGGAAGTTTGAACAGTCAACAGAGAAGAACACCTAGGAAAATT 1198
Db 1028 -----LeuLysAsnLysHisGluSerMetIleSerGluLeuGluValArgLeu 1043
QY 1199 TTGAGGCTCACAAGAAACATCTGAGAAATTTTCATGGCGCAGCAAAAGAAAGATCTAGG 1258
Db 1044 LysLysGluLysSerArgGlnGlu-----LeuGluLysLeuLysArg 1058
QY 1259 AAGATCATCGGAGGAGAAAGAAACATCTGTAAAG---ACTGAATCGTGGCAGGAGTA 1315
Db 1059 LysMetAspGlyGluAlaSerAspLeuHisGluGlnIleAlaAspLeuGlnAlaGlnIle 1078
QY 1316 ACACCTAATAAACTGAAGTTTGGAAAAGGAACATCTAATATGATGATGCTCTTACATA 1375
Db 1079 AlaGluLysMetGlnLeuAlaLysLysGluGluGluLeuGlnAlaLeuAlaArg 1098
QY 1376 AAAGAAACATCTCAAAAGCAAGTACAAATGTGATGTGATGCTGTAGAGCCTATATTC 1435
Db 1099 LeuGluAspGluThrSerGlnLysAsnAsn----- 1108
QY 1436 AGTCTTTTGGCACACCGACTATTGAAAATTCACAGTGTACAAAGTTGAGGAAGACTTT 1495
Db 1109 AlaLeuLysLysIleArgGluLeuGlu---GlyHisIleSerAspLeuGlnGluAspLeu 1127
QY 1496 AATCTTCTACCAAGATTATCTCTAAGAGTGTCTGACAGATATTATACGCTGTTACCTGAT 1555
Db 1128 Asp-----SerGluArgAlaAlaArgAsnLysAlaGluLysGlnLys 1141
QY 1556 GCTACATATCAAAAGATATCAAAACAAATAAATACAAAATAGAAAGTTCAGATGTTCCCA 1615
Db 1142 ArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAspThrLeuAspThr 1161
QY 1616 TCAGAAATCCAAACAGAG-----GAAGATGAAGAATATCTTGGGATCTCGG 1663
Db 1162 ThrAlaThrGlnGlnGluLeuArgAlaLysArgGlnGlnValThrValLeuLysLys 1181
QY 1664 AGTCTCTTTCAGAGTTCTGCAAGACTCAAGTGTGTATACCTCAGTCTATGATTCAGAAA 1723
Db 1182 AlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGlu---MetArgGlnLys 1200
QY 1724 GTAATGGAGATAATAGAGAAAGTAGAGACTCTCTGAGAAAGCCATCTGCTCTCAAGCCT 1783
Db 1201 HisThrGlnVal-----ValGluGluLeuThrGluGlnLeuGluGlnPheLysArg 1217

QY	1784	GCCGTN-----GAAATGCAAAAGACGTGTTCCAAATAAAGCCCTTGAAATTGAAG	1831
DB	1218	AlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLysGluAsnAlaAspLeuAla	1237
QY	1832	AATGAACAAACATTGAGACGAGCTCAGATGTTCCCATCAGAAATCCAAACAAAGAGCAGAT	1891
DB	1238	GlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLysLeu	1257
QY	1892	GAAGAAAAATCTTGGGATCTCGAGAGCCCTGTGAGACGGCTTTCACAGAAGGATGTGTAT	1951
DB	1258	GluValGlnLeuGlnGluLeuGlnSerLysCysSerAspGlyGluArgAlaAspAlaGlu	1277
QY	1952	TTACCCAAAGCTACACAT-----CAAAAAGATTCGATACCTTAAGTGGAAATTAGAA	2005
DB	1278	LeuAsnAspLysValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeuSer	1297
QY	2006	GAGTCTCCTGTTAAAGATGCTCTTCTGAAGCCTCCTGTGCAAGGAAAGTTTCTCTTCCA	2065
DB	1298	GluAla-----Glu	1300
QY	2066	AATAAGCCTTAGAATTA---AGGACAGAGAAACATTCAACAGCAGAGCTCTCTGTATAA	2122
DB	1301	GlyLysAlaIleLysLeuAlaLysGluValAlaSerLeuGlySerGlnLeuGlnAspThr	1320
QY	2123	GATGGTCTCTGGAAGCCTACCTGTGGAGGAAAGTTTCTCTCCAAATAAGCCTTAGAA	2182
DB	1321	GlnGluLeuLeuGlnGluThrArgGlnLysLeuAsnValSerThrLysLeuArgGln	1340
QY	2183	TTAAGGACAGAGAAACACTCAAAGCAGAGTCTCTGATTAATGATGTCTTCTGAGCCCT	2242
DB	1341	LeuGluAspGluArgAsnSerLeuGlnGluGluLeuAspGluGluMetGluAlaLysGln	1360
QY	2243	ACCTGTGGAAGGAAAGTTTCTCTCCAAATAAAGCTTTAGAAATTGAAGGACAGAGAAACA	2302
DB	1361	AsnLeuGluArgHisIleSerThrLeuAsnIleGlnLeu-----	1373
QY	2303	TTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACAAAGAGATGATGAAGAAATCT	2362
DB	1374	-----SerAspSerLysLysLys-----Leu	1380
QY	2363	TGGATTTTCAGAGTTTCTCTTGAGACTCTCTTACAGAATCATGTGTGTTTACCACAGGCT	2422
DB	1381	GlnAspPheAlaSerThrValGluSerLeuGluGluGly-----LysLys	1395
QY	2423	ACACATCAAAAAGAAATTCGATACCTTAAGTGGGAAATTAGAAGAG---TCTCTGTATAA	2479
DB	1396	ArgPheGlnLysGluIleGluSerLeuThrGlnGlnItyrGluGluLysAlaAlaIaTyr	1415
QY	2480	GATGGTCTCTGGAAGCCTACCTGTGGGAATCAAAATTTCTCTCCAAATAAGCCTTAGAA	2539
DB	1416	AspLysLeuGluLysThrLysAsnArgLeuGlnGlnLeuLeuAspSerLeuValValAsp	1435
QY	2540	TTGAAGGACAGAGAAACATTCAAAGCAGAGATGTGAGTCTCTGTAGAGTCCACATTCAGT	2599
DB	1436	LeuAspAsnGlnArgGlnLeu-----ValSerAsnLeuGluLysLysGlnLys	1451
QY	2600	CTTTTTCGCAACCGACTACTCGAAATTCACAG--TCTACAAAGTTGAGGAGAGAC---	2653
DB	1452	LysPheAspGlnLeuLeuAlaGluGluLysAsnIleSerSerLysTyrAlaAspGluArg	1471
QY	2654	-----TTTAATCTTACTACAGGAGGGAGCAACAAAGACAGTAATCGGCAACACAGAA	2707
DB	1472	AspArgAlaGluAlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAla	1491
QY	2708	CGTGATATTGGCAATTATTGAACGAGCTCCACAGACTCAAAACAAATAAGATGCCACATCA	2767
DB	1492	LeuGluGluAlaLeuGluAlaLysGluGluLeuGluArgThrAsnLysMetLeuLysAla	1511
QY	2768	GAATTAGGAAGAAAGAGATACAAATCAACTTCAGATTCGTGAGATTCTCTGTGAGT	2827
DB	1512	GluMet-----GluAspLeuValSerSerLysAspValGlyLysAsnValHis	1528

Qy	2828	GATACACAGAAATTATGAGTGTCTTTACCTGTGAGGCTACATATCAAAAGAGAAATAAGACA	---	2884
Db	1529	GlueuGlulysSerLysArgAlaLeuGlu- --ThrGlnMetGluGluMetLysThrGln	---	1547
Qy	2884	-----	-----	2884
Db	1548	LeuGluGluLeuGluaspGluLeuGlnAlaThrGluaspAlaLysLeuArgLeuGluVal	1567	
Qy	2885	-----	-----ACAAATGCCAAATATAGAGAGTCT	2908
Db	1568	AsnMetGlnAlaLeuLysValGlnPheGluArgAspLeuGlnAlaArgAspGluGlnAsn	1587	
Qy	2909	CCTGAAAGACCTTCTCATT- --GAGCCTGCCCTGAAATGCCAAAC	2953	
Db	1588	GlueuLysArgGlnLeuGlnArgGlnLeuHisGluPyrGluThrGluLeuGluasp	1607	
Qy	2954	TCTGTTCCAAATAAAGGCTTAGAATGGAAGAAATAACAACATTGAGAGAGATTCAACT	3013	
Db	1608	GluArgLysGlnArgAlaLeuAlaAlaAlaLysLysLysLeuGluGlyAspLeuLys	1627	
Qy	3014	ACCTATCAAAAATCTTGGATGCACCTCTCTGTGAAAGAGACGGCACTTAAAAAA	3073	
Db	1628	AspLeuGluLeuGlnAlaAsp- --SerAlaIleLysGlyArgGlu- --	1641	
Qy	3074	GATAACTGTCAACAAATTTACAGCAAAATGCAACAAATGAAATAGTTTGTGTACTA	3133	
Db	1642	GluAlaIleLysGlnLeuLysLeuGlnAlaGlnMetLysasp- --Phe	1657	
Qy	3134	CAAAAGGAAGCTGTGAGAGCGAAGAAATAAATTCACAGTTA- --GAG	3178	
Db	1658	GlnArgGluLeuGluaspAlaArgAlaSerArgAspGluIlePheAlaThrAlaLysGlu	1677	
Qy	3179	AACCAAAA- --GCTAAATGGCAACAGAGCTCTGCAGTGTGACAGTGCCTTTAAAT	3232	
Db	1678	AsnGluLysLysAlaLysSerLeuGluAlaAspLeuMetGlnLeuGlnLysLeuAla	1697	
Qy	3233	CAAGAAGAGAGAGAGAAATGTGATATATTAAGAAAGAAATATAGACCCGAA- --	3289	
Db	1698	AlaAlaGluArgAlaArgLysGlnAlaAspLeuLysGluGluLeuAlaGluGluLeu	1717	
Qy	3290	-----	-----GAGCAACTTAGGAAAAAGTTAGAGTGAAA	3319
Db	1718	AlaSerSerLeuSerGlyArgAsnAlaLeuGlnAspGluLysArgArgGluGluAlaArg	1737	
Qy	3320	---CACCAACTTGAACAGACTCTCAGAATACAAGATATAGAAATGAAAAGTGTA- --	3370	
Db	1738	IleAlaGlnLeuGluGluLeuGluGluGluGluGluGluGlnGlyAsnMetGluAlaMetSerAsp	1757	
Qy	3371	-----ACAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAGTGAAT	3415	
Db	1758	ArgValArgLysAlaThrGlnGlnAlaGluGlnLeuSerAsnGluLeuAlaThrGluArg	1777	
Qy	3416	GATCTCTTTCATCAAAAT- --	3433	
Db	1778	SerThrAlaGlnLysAsnGluSerAlaArgGlnGlnLeuGluArgGlnAsnLysGluLeu	1797	
Qy	3434	-----	-----TGCATGTTGAAAAAGGAAATTGCC	3457
Db	1798	LysSerLysLeuGlnGluMetGluGlyAlaValLysSerLysPheLysSerThrIleAla	1817	
Qy	3458	ATGCTAAAACCTGGAAGTAGCCACACTCAAAATCATCAACACAGGCTGAAGGAAAAATAATAC	3517	
Db	1818	AlaLeuGluAlaLysIleAlaGlnLeuGluGluGlnValGlnGlnGlnAlaArgGluLys	1837	
Qy	3518	TTTGAGGACATTAAAGATTTTACAGAAGAAAGATGCTGAATCTCAAAATGACCCCTTAAACATG	3577	
Db	1838	GlnAlaAlaAlaLysAlaLeuLysGlnArgAspLysLysLysLysGluMetLeu- --Leu	1856	
Qy	3578	AAACAGAAAACAGTAAACAAAAGGCGATCTCAGTATAGAGACAGCTTAAAGTCTCAGC	3637	
Db	1857	GlnValGluaspGluArgLysMetAlaGluGlnIleLysGluGlnAlaGluLysGlyAsn	1876	
Qy	3638	GCAGAGAACACGATGCTGACTTCTTAAATTTGAAGGAA	3673	

QY	1436	AGTC	TTTTGGCACCGGACTATTGAAATTC	-----CAGTGT	1474
Db	683	AsnVal	LeuIysProAlaAspAspLysSerLysSerAspValThrAspLysSer		702
QY	1475	ACAAAG	TTCGAGCAACTTAACTTCCTGCTACCAAGATTATCTCTAAGAGTGCTGCACAC	1534	
Db	703	LysLys	ThrThrGluAsp-----GlnThrLysValAlaThrAspSerLysLeuGlu	719	
QY	1535	AATTAT	ACGTGTTTACCTGTGATCATATCAAAAAGATATCAAAAACAATAAATACAAA	1594	
Db	720	Lys-----	AlaAlaAspThrThrLysGlnIleGluThrGluThrVal-----	733	
QY	1595	ATAGAAG	ATCAGATGTTCCCATCAGATCCAAACGAGAGGAGAGATGAAGATATTCCTGG	1654	
Db	734	ValAsp	Asp-----LysSerLysLysValLeuLysLysThrGlu	748	
QY	1655	GATTC	GGGAGTCTCTTTGAGCTTCTGCAAGAGCTCAAGTGTGTAT-----	1702	
Db	749	LysSerAsp	SerPheIleSerGlnLysSerGluThrProValValGluProThrLys	768	
QY	1703	---CCTG	AGTCTATGATCAGAAAGTAATGGAGATAATAGAGACTAGAGAGCTTCCT	1759	
Db	769	ProAla	GluSerGluAlaGlnLysIleAlaGluValAsnLysAlaLysLysGlnLysGlu	788	
QY	1760	GAGAAG	CCATCTGCCTTCAAGCCCTGCGGTGAAATGCAAAAGAGCTGTTCCAAATAAAGCC	1819	
Db	789	ValAsp	AsnLeuLysArgGluAlaGluValAlaAlaLysLysIleAlaAspGluLys	808	
QY	1820	TTTCA	ATTGAAGTATGACAAACATTGACAGCAGCTCAGATGTTCCCATCAGATCCAAA	1879	
Db	809	LeuLys	IleGluAlaGluAlaAsnIleLysLysThrAlaGluValGluAlaAlaLysLys	828	
QY	1880	CAAAA	GGAC--GATGAGAAATTCCTGGATTCCTGAG-----AGTCCC	1921	
Db	829	GlnLys	GlnLysAspGluGlnLeuLysLeuGluThrGluValValSerLysLysSerAla	848	
QY	1922	TGTGAG	ACGTTTCA--CAGAAGATGCTGTTATTTACCAAGACTACACAT-----	1969	
Db	849	AlaGlu	LysLeuGluLysGlnAlaGlnIleLysLysAlaGluAlaAspAla	868	
QY	1970	-----	CAAAAAGATTCGAT-----	1984	
Db	869	ValLys	LysGlnLysGluLeuAsnGluLysAsnLysLeuGluAlaLysLysSerAla	888	
QY	1985	-----	-----ACCTTAAGTGAAATTAGAAGAGTCT	2011	
Db	889	AlaAsp	LysLeuLysGluGluGluSerAlaAlaLysSerLysLysValSerGluGlu	908	
QY	2012	CCTGT	TAAAGATGGTCTCTGAAGCCTACC--TGTGGAAGGAAGTT-----	2056	
Db	909	SerVal	LysPheGlyGluGluLysLysThrLysAlaGlyGlnLysThrValGlnValGlu	928	
QY	2057	TCCTCT	TCCA-----AATAAGCCTTAGAATTAAGACAGAGAAAACATTCAAGACA--	2107	
Db	929	SerGlu	ProThrSerLysLysThrIleAspThrLysAspValGlyAlaThrGluProAla	948	
QY	2108	---GAGT	CTCCTGATAAAGATGGTCTTCTGAAGCTACCTGTGGAAGGAAGTTTCTCTT	2164	
Db	949	AspGlu	ThrProLysLysLysIleIleLysLysLysThrGlnLysSerAspSerIle	968	
QY	2165	CCAAA	TAAAGCCTTAGAATTAAGACAGAGAAACACTCAAAGCAGAGTCTCCTGATAAT	2224	
Db	969	SerGln	LysSer-----AlaThrAspSerGlnLysVal--SerLysGlnLysGluGln	985	
QY	2225	GATGGT	TTCTGAAGCCTACCTGTGGAAGGAAGTTTCTTCTCCAAATAAAGCTTTAGAA	2284	
Db	986	AspGlu	ProThrLysProAlaValSerGluThrGlnMetValThrGluAlaAspLysSer	1005	
QY	2285	TTGA	GCACAGAAACA-----TTCAAAGCAGCTCAGATGTTCCCATCAGAAATCC	2335	
Db	1006	LysLys	GlnLysGluThrAspGluLysLeuLysLeuAspAlaGluIleAlaLysThr	1025	

Qy	2336	AAACAAAGAGTATGATGAAGAAAATCTCTGGGATTTTTCCTGAGAGTTTCTCTTGAGACTCTCTCTTA	2399
Db	1026	LysGlnGluAlaAspGluLysSerLysLeuAspAlaGlnGluLysIleLysLysValSer	1045
Qy	2396	CAGAAATGATGTGTCTTTTACCACAG-	2419
Db	1046	GluAspAlaAlaAaRgysGluLysGluLeuAsnAspLysLeuLysLeuGluSerGlu	1065
Qy	2420	---GCTACACATCAAAAGAATTTCGATACCTTAAGTGGAAAATTTAGAGAG-	2467
Db	1066	IleAlaThrLysLysAlaSerAlaAspLysLeu-----LysLeuGluGluGlnAlaGln	1083
Qy	2468	-----TCTCCTGATATAAGATGGCTTCTCTG	2491
Db	1084	AlaLysLysAlaAlaGluValGluAlaAlaLysLysGlnLysGluLysAspGluGlnLeu	1103
Qy	2492	AAG-----CCTACCTGTGGAATGAAATTTCTCTTCCAAATAAGCCTTAGAATTG	2542
Db	1104	LysLeuAspThrGluAlaAlaSerLysLysAlaAlaAlaGluLysLeuGluLeuGlu---	1122
Qy	2543	AAGCAGACAGAAACATTCAAAGCAGAGATGTGAGTTCTGTAGAGTCCACATTCAGTCTT	2602
Db	1123	LysGlnAlaGlnIleLysLysAlaAlaGlyAlaAspAlaValLysLysGlnLysGluLeu	1142
Qy	2603	TTTGGCAAAACCGACTACTGAA---AATTCAGTCTACA-----AAAGTTTGAG	2647
Db	1143	AspGluLysAsnLysLeuGluAlaAsnLysLysSerAlaAlaGlyLysLeuLysIleGlu	1162
Qy	2648	GAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACCTGCACAACAGGAA	2707
Db	1163	GluGluSerAlaAlaLysSerLysGln-----	1171
Qy	2708	CGTGATATTGGCATTTATTGAACGAGCTCCACAGATCAACAAATAAGATGCCACATCA	2767
Db	1172	-----ThrValGluGluGlnAlaLysLeuAspAlaGlnThrLysAlaLysThrAla	1188
Qy	2768	GAATTAGGAAGAAAA-----GAAGATACAAAATCAACTTCAGATTCTGAGATTATCTCT	2821
Db	1189	GluLysGlnThrLysLeuGluLysAspGluLysSerThrLysGluSerGluSerLysGlu	1208
Qy	2822	GTGAGTATACACAGATTATGAGTGTTTACCTGAGCTCATATCAAAAGAATAAAG	2881
Db	1209	ThrValAspGluLysProLysLysValLeuLysLysLysThrGluLysSerAspSer	1228
Qy	2882	ACAACAAATGGCAAAATAGAAGACTTCCT-----GAAAGGCCCTTCTCACTTTGAG	2932
Db	1229	SerIleSerGlnLysSerGluThrSerLysThrValValGluSerAlaGlyProSerGlu	1248
Qy	2933	CTGCGCCTACTGAATGCAAAACTGTGTCCAAATAAGGCTTAGAATGGAGAATAACAA	2992
Db	1249	SerGluThrGlnLysValAlaAspAlaAlaArgLysGlnLysGluThrAspGluLysGln	1268
Qy	2993	ACATTGAGACGACATTCAACTACCTATCAAAATCTTGGATGACCTTCCTCTCTGTGAA	3052
Db	1269	LysLeuGluAlaGlu---IleThrAlaLysLysSerAlaAspGluLysSerLysLeuGlu	1287
Qy	3053	AGAGGAAGGAACTTAAAAAGATAACTGTGCAACAAATTACAGCAAAATATGGACAAATG	3112
Db	1288	AlaGluSerLysLeuLysLys-----AlaAlaGluValGluAlaAla	1301
Qy	3113	AAAAATAAGTTTCTGTGACTACAAAGGAACGTGCAGAACGAAAGAAATAAAATCACAG	3172
Db	1302	LysLys-----GlnLysGluLysAspGluGlnLysLeuLysAspThrGlu	1316
Qy	3173	TTAGAACCAAAAAGCTAAATGGCAACAGAGCTCTGCAGTGTGAGATTGCCCTTTAAAT	3232
Db	1317	AlaAlaSerLysLysAlaAlaAlaGluLys-----LeuGluLeuGluLysGln	1332
Qy	3233	CAAGAAGACGAGAGAGAAAGAAATGTCGATATTAAAGAAAAAAATAGACCGAGAG	3292
Db	1333	SerHisLysLysAlaAlaGluValAspAlaValLysLysGlnLysGluLeuGluGlu	1352
Qy	3293	CAACTTAGG-----AAAAAGTTAGAGTG	3316

```

Db 1353 LysGlnArgLeuGluSerGluAlaAlaThrLysLysAlaAspAlaGluLysLeuLysLeu 1372
      :: ||| ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3317 AAACACCAACTTGAACACAGCTCTCAGCAATACAGATATAGAANTTGAAGTGAACAAAGT 3376
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1373 GluGluGlnLysLysLysAlaAlaGluAlaLeuIleGluLeGlnLysGluGlnGlu 1392
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3377 AATTGAATCAGGTTCTTCACATCATCAGAAAGTGAATGAAATCATCTCTTCATGAAAATTGC 3436
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1393 LysLeuAlaGlnGluGlnSerArgLeuGluAspGlu 1404
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3437 ATGTTGAAAAGAAATTCGCATGCTAAAAGTGGAGTACCCACATCAACATCAACAC 3496
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1405 --AlaLysLysSerAlaGluLysGlnLysLeuGluSerGluThrLysSerLysGlnThr 1423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3497 CAGGTGAAGGAAATAAATCTTTGGAGC-----ATTAGATTTTACAAGAAAAG 3547
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1424 GluGluAlaProLysGluSerValAspGluLysProLysLysValLeuLysLysLys 1443
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3548 AATGCTCAACTTCAAATGACCTTAAACTGAAACAGAAAAACA 3589
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1444 ThrGluLysSerAspSerSerLysSerGlnLysSerLysSer 1457
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
Tl3564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: Tl3564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: Tl3564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Alignment Scores:
Pred. No.: 7,65e-10 Length: 5327
Score: 312.50 Matches: 228
Percent Similarity: 34.80% Conservative: 167
Best Local Similarity: 20.09% Mismatches: 485
Query Match: 4.85% Indels: 255
DB: 2 Gaps: 39

US-09-602-362E-26 (1-3673) x Tl3564 (1-5327)
Qy 794 AAATGCACGCCCTCATGCTTGCCATATGTGAAGGCTCATCAGAGTAGTCGCATGCTT 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2984 LysThrThrSerArgValSerValAlaAspSerLysAspGluLysSerLeu 3003
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 854 CTTCCAGCAAAATGTT-----GAGCTTTTGGCTGAAGACATCATGGAATAACTGCA 904
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3004 ValSerGlnGluAlaSerArgProGluSerGluAlaGluSerLeuLysAspAlaAla 3023
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 905 GAAGTTATCTGCTGCTGCTGAGTGAATTAATCATTCATCAACAACTTTTGGACATATA 964
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3024 ProSerGlnGluThrSerArg----- 3030
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 965 CGAAATTAATCTAAAATCTCAAATATCCAAATCCAGAGGAACATCTACAGGAACACCT 1024
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3031 -----ProGluSerValThrGluSerValLysAspGluLysSerProValAlaSer 3047
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1025 GATGAGGCT-----GCACCTTGGCGGAAAGAACACCTGACACGGCTGAAGCTTG 1075
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 3048 LysGluAlaSerArgProAlaSerValAlaGluAAsnAlaLysAspSerAlaAspGluSer 3067
QY 1076 CTGGAAACACCTGACGAGGCTGCACGCTGGTGGAGGAGCGTCTCCCAAAATTCAA 1135
Db 3068 LysGluGlnArgProGlnSerLeuProGlnSerLysAlaGly--SerIleLysAspGlu 3086
QY 1136 TGCTCGGGAAGCAACATCTCGAAAGTTTGACAGCTCAACAGAGAAACACCTAGGAAA 1195
Db 3087 LysSerProLeuAlaSerLysAspGluAlaGluLysSerLysGluGluSerArgArg-- 3105
QY 1196 ATTTGAGCCCTACAAAAGAACATCTGAGAAATTTTCATGCCAGCAAAAGAGATCT 1255
Db 3106 -----GluSerValAlaGluGlnPheProLeuValSerLysGluValSer 3120
QY 1256 AGGAGATCACATGGGAGGAGAAAAGAACATCTGTAAGACTGAATGC----- 1303
Db 3121 ArgProAlaSerValAlaGlu-----SerValLysAspGluAlaGluLysSerLys 3137
QY 1304 -----GTGCACGA----- 1312
Db 3138 GluGluSerProLeuMetSerLysGluAlaSerArgProAlaSerValAlaGluSerVal 3157
QY 1313 -----GTAACACCTTAATAAAGCTGAAGTTTGGAAAAAGAACATCTAATATGATT 1363
Db 3158 LysAspGluAlaGluLysSerLysGluGluSerArgArgGluSerValAlaGluLysSer 3177
QY 1364 GCATGCTCTACAAAAGAACATCTACAAAAGCAAGTACAAATGTGGATGTGAGTCTGTA 1423
Db 3178 ProLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerValLysAspGlu 3197
QY 1424 GAGCCTATATTCAGTCTTTTGGCACACGAGCTATTGAAAATTCACAGGTACAAAAGTT 1483
Db 3198 -----AlaAspLysSerLysGluGluSerArgArgGluSerGly 3210
QY 1484 GAGGAGAGCTTTAATCTGTACCAAGATTATCTTAAGAGTGCTGCACAGAAATTATACG 1543
Db 3211 AlaGluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSer 3230
QY 1544 TGTTTACCTGATCTACATATCAAAAAGATATCAAAAATCAATAATCAAAAATAGAGAT 1603
Db 3231 IleLysAspGluAlaGluLysSerLysGluGluSerArgArgGluSerValAlaGluLys 3250
QY 1604 CAGATGTTCCATCA----- 1618
Db 3251 SerProLeuProSerLysGluAlaSerArgProThrSerValAlaLysSerValLysAsp 3270
QY 1619 -----GAATCCAAAAGAGAGGAGATGAAGATATTTGGGATTCGGGAGTCTC 1669
Db 3271 GluAlaGluLysSerLysGluGluSerArgAspSerValAlaGluLysSerProLeu 3290
QY 1670 TTTGAGAGTCTCBAAGACTCAAGTGTATACCTGAGTCTATGAT-----CAG 1720
Db 3291 AlaSerLysGluAlaSerArgProAlaSerValAlaGluSerValGlnAspGluAlaGlu 3310
QY 1721 AAAGTAATGGAGATAAATAGAAAGTAGAAGAGCTTCTCGAAGGCCATCTGCTTCAAG 1780
Db 3311 LysSerLysGluGluSerArgArgGluSerValAlaGluLysSerProLeuAlaLysLys 3330
QY 1781 CTGTC-----GTNGAAATGCAAAAGACTGTTCCAAAATAAGCCCTTGAAATGAG 1831
Db 3331 GluAlaSerArgProAlaSerValAlaGluSerIleLysAspGluAlaGluLysSerLys 3350
QY 1832 AATGAACAAACATTGAGACGA-----GCTCAGATGTTCCCATCAGAATCCAAA----- 1879
Db 3351 GluGluSerArgGluSerValAlaGluLysSerProLeuAlaSerLysGluAlaSer 3370
QY 1880 -----CAAGACGATGAGAAAT 1900
Db 3371 ArgProThrSerValAlaGluSerValLysAspGluAlaGluLysSerLysGluGluSer 3390
QY 1901 TCTTGGGATCTCAGAGTCCCTGTGAGCGGTTTCACAGAGGATGTGATTATACC----- 1957
Db 3391 SerArgAspSerValAlaGluLysSerProLeuAlaSerLysGluAlaSerArgProAla 3410

QY 1958 -----AAAGCTACACATCAAAAAGAAATTCGATACCTTAAGT 1993
Db 3411 SerValAlaGluSerValGlnAspGluAlaGluLysSerLysGluGluSerArgGlu 3430
QY 1994 GGAAATTTAGAGAGTCTCTCT----- 2014
Db 3431 SerValAlaGluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAla 3450
QY 2015 -----GTTAAAGATGCTCTCTCTCAAG----- 2035
Db 3451 GluSerValLysAspAspAlaGluLysSerLysGluGluSerArgArgGluSerValAla 3470
QY 2036 -----CCTACTGTGGAAGAAAGTTTCTTCCAAATAAAGCCTTAGAA--TTA 2083
Db 3471 GluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerVal 3490
QY 2084 AAGCACAGAGAAACATTCAAGACAGAGTCTCTCTGATAAGATGCTCTCTCTGAAG----- 2137
Db 3491 LysAspGluAlaGluLysSerLysGluGluSerArgArgGluSerValAlaGluLysSer 3510
QY 2138 CCTACTGTGGAAGAAAGTTTCTTCCAAATAAAGCCTTAGAAATTA----- 2185
Db 3511 ProLeuProSerLysGluAlaSerArgProThrSerValAlaGluSerValLysAspGlu 3530
QY 2186 -----AAGCACAGAGAAACACTCAAGCAGAGTCT---CCTGATAATGATGCTCTCTG 2236
Db 3531 AlaGluLysSerLysGluGluSerArgArgGluSerValAlaGluLysSerSerLeuAla 3550
QY 2237 AAGCCTACTGTGGAAGAAAGTTTCTCTT-----CCAAATAAA 2275
Db 3551 SerLysGluAlaSerArgProAlaSerValAlaGluSerValLysAspGluAlaGluLys 3570
QY 2276 GCTTTAGAAATGAAGGACAGAGAAACATTCAAGCAGCTCAGATGTTCCCATCAGATCC 2335
Db 3571 SerLysGluGluSerArgGluSer-----ValAlaGluLysSerProLeuAlaSer 3588
QY 2336 AAACAA-----AAGGATGATGAAGAAATTTCT 2362
Db 3589 LysGluAlaSerArgProAlaSerValAlaGluSerValLysAspGluAlaGluLysSer 3608
QY 2363 TGGGATTT-----GAGAGTTCTCTTGAG-----ACTCTCTACAGAAATGATGTG 2407
Db 3609 LysGluValSerArgArgGluSerValAlaGluLysSerProLeuProSerLysGluAla 3628
QY 2408 TGTTTACCAGGCTACACATCAAAAAGAAATTCATACCTTAAGTGGAAATTAAGAGAG 2467
Db 3629 SerArgProThrSerValAlaGluSerValLysAspGluAlaAspLysSerLysGluGlu 3648
QY 2468 TCTCTGATAAAGATGCTCTCTCAAG---CCTACCTGTGGAATGAAATTTCTCTTCCA 2524
Db 3649 SerArgArgGluSerGlyAlaGluLysSerProLeuAlaSerMetGluAlaSerArgPro 3668
QY 2525 AATTAAGCCTTAGAATTTGAAGACAGAGAAACATTCAAAGCAGAGGATGTGAGTCTGTA 2584
Db 3669 ThrSerValAlaGluSerValLysAspGluThrGluLysSerLysGluGluSerArgArg 3698
QY 2585 GAGTCCACATTCAGTCTTTTGGCAACCGACTACTGAAAATTCACAGTCTACAAAAGTT 2644
Db 3689 GluSerValThrGluLysSerProLeuProSerLysGluAlaSerArgProThrSerVal 3708
QY 2645 GAGGAAGACTTTAATCTTACTACCAAGAGGAGGAGCAACAAAGACAGTAAGTGGACACAG 2704
Db 3709 AlaGlu-----SerValLysAspGluAlaGluLys-----SerLysGluGlu 3722
QY 2705 GAACGTGATATTGGCATTATTGAACGAGCTCCCAAGATCAACAAATAAGATGCCACA 2764
Db 3723 SerArgArgGluSerValAlaGluLysSerPro----- 3733
QY 2765 TCAGAAATTTAGGAAGAAAGAGATACAAAATCAACTTCAGATTCAGATATTC----- 2818
Db 3734 -----LeuAlaSerLysGluSerSerArgProAlaSerValAlaGluSerIleLysAsp 3751

QY 2819 -----TCTGTGAGTGTATACAGAAATATGAGTGTCTTACTGAGGCTACATATCAA 2869
 Db 3752 GluAlaGluGlyThrLysGlnGluSerArgGluSerMetProGluSerGlyLysAla 3771
 QY 2870 AAGAAATAAAGCAACAAATGCGAAATAA---GAAGAGTCTCTGAAAGGCTTCTCAC 2926
 Db 3772 GluSerIleLysGlyAspGlnSerSerLeuAlaSerLysGluThrSerArgProAspSer 3791
 QY 2927 TTGAGCCTCCACTGAAATGCAAACTCTGTTCCAAATAAAGGCTTAGAATGGAAGAT 2986
 Db 3792 ValValGluSerValLysAspGluThrGluLysProGluSerAlaIleAspLysSer 3811
 QY 2987 AACAAACATTTGAGAGCAGATTCACCTACCTCAATCAAAAATCTTGATGCACTTCTTCT 3046
 Db 3812 GlnValAlaSerArgProGluSerValAlaValSer----- 3823
 QY 3047 TGTGAAGAGGAGGAACTTAAAAAGATACTGTGAACAAATTACAGCAAAATGGAA 3106
 Db 3824 -----AlaLysAspGluLysSerProLeuHisSerArgProGlu 3836
 QY 3107 CAATGAAATAATAGTTTGTGTACTACAAAGCAACTGTCAAGCGGAAAGAAATAAA 3166
 Db 3837 SerValAlaAspLysSerProAspAlaSerLysGluAlaSerArgSerLeuSerValAla 3856
 QY 3167 TCACAGTTAGAGAACCAAAAGCTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTGCCT 3226
 Db 3857 GluThrAlaSerProIleGluGluGlyProArgSerIleAlaAspLeuSerLeuPro 3876
 QY 3227 TTAATCAAGAGAGAGAGAGAGAAATGCGATATATTAATAAGAAATAATAGACCC 3286
 Db 3877 LeuAsnLeuThrGlyGluAlaLysGlyLysLeuProThrLeuSerSerProIleAspVal 3896
 QY 3287 GAAGAGCAACTTAGGAAAAAGTTAGAGTGAAGTGAACACACCACTT----- 3328
 Db 3897 AlaGlu-----GlyaspPheLeuGluValLysAlaGluSerSerProArgProAlaVal 3914
 QY 3328 ----- 3328
 Db 3915 LeuSerLysProAlaGluPheSerGlnProAspThrGlyHisThrAlaSerThrProVal 3934
 QY 3329 ---GACAGACTCTCAGATACAGATATAGATTG-----AAAAGTGAACAGTAAT 3379
 Db 3935 AspGluAlaSerProValLeuGluGluGluLeuValValGluGlnHisThrSerGly 3954
 QY 3380 TTGAATCAGGTTTCTCACACTCATGAAAGTGAA-----AATGATCTCTTTCAT 3427
 Db 3955 ValGlyAlaThrGlyAlaThrAlaGluThrAspLeuLeuAspLeuThrGluThrLysSer 3974
 QY 3428 GAAATTCATGTTGMAAAGGAAATGCCATGCTAAACTGAACTGACCTGACCTGAA 3487
 Db 3975 GluThrValThrLysGlnSerGluThrThrLeuPheGluThrLeuThrSerLysValGlu 3994
 QY 3488 CATCAACACACAGCTGGAAGAAAT-----AAATACTTTGAGGACATTAAGATTTTACAA 3541
 Db 3995 SerLysValGluValLeuGluSerSerValLysGlnValGluLysValGlnThrSer 4014
 QY 3542 GAAAGAAATCTGAACTTCAATGACCTTAAACTGAAACAGAAAAACATACAAAAGG 3601
 Db 4015 ValLysGlnAlaGluThrThrValThrAspSerLeuGluGln-----LeuThrLysLys 4032
 QY 3602 GCATCTCAGTATAGAGAGAGCTTAAAGTTCTGACGCGCAGAGAAC 3646
 Db 4033 SerSerGluGlnLeuThrGluIleLysSerValLeuAspThrAsn 4047

RESULT 12

A56539

giantin - human

N;Alternate names: macrogolin

C;Species: Homo sapiens (man)

C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999

R;Accession: A56539; S37536

R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994

A>Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (C)
 A;Reference number: A56539; MUID:94187728; PMID:7511208
 A;Accession: A56539
 A;Molecule type: mRNA
 A;Residues: 1-3259 <SSE>
 A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
 C;Genetics:
 A;Gene: GDB:COLGB1; GCP; GCP371
 A;Cross-references: GDS:454958
 A;Map position: 3q13.31-3q13.31
 C;Superfamily: giantin
 C;Keywords: coiled coil; Golgi apparatus; transmembrane protein
 F:3238-3254/Domain: transmembrane #status predicted <TMN>

Alignment Scores:
 Pred. No.: 8.77e-10 Length: 3259
 Score: 311.00 Matches: 235
 Percent Similarity: 36.70% Conservative: 212
 Best Local Similarity: 19.29% Mismatches: 425
 Query Match: 4.82% Indels: 346
 DB: 1 Gaps: 48

US-09-602-362E-26 (1-3673) x A56539 (1-3259)

QY 680 GTGCAAAACAGGCTAGCTCCTCACACCCCTTTTACTGCCATACAGAAAAGCAAGCAA 739
 Db 1101 LeuGlnAspLysThrAsnGlnIleAspLeuLeuGlnAlaGluIleSerGluAsnGlnAla 1120
 QY 740 ACTGTGGAATTTTACTAACAAAATGCAAAATGCAAAACGCAATTAATGAG----- 790
 Db 1121 IleIleGlnLysLeuIleThrSerAsnThrAspAlaSerAspGlyAspSerValAlaLeu 1140
 QY 791 TCTAAATGTCAGAGCCCTCATGCTTGCATATGTGAAGGCTCATCAGACATAGTCGGCATG 850
 Db 1141 ValLysGluThrValValIleSerProCysThrGlySerSerGluHisTrpLysPro 1160
 QY 851 CTTCTTCAGCAAAATGTT-----GAGCTCTTCTGCTGAAGACATACATGAATTAAT 901
 Db 1161 GluLeuGluGluLysIleLeuAlaLeuGluLysGluLysGluLysLysLeu 1180
 QY 902 GCAGACGTTATGCTGCTGCTCGGGAGTTAATTTACATTCATCAACATTTTGGAAAT 961
 Db 1181 GlnGluAlaLeuThrSerArgLysAlaIleLeuLysLysLysLysLysLysLys 1200
 QY 962 ATACGA-----AAATACCTTAAATTCCTCAAAATACCAATCCAGAAAGCAATCT 1012
 Db 1201 LeuArgGluGluLysGlnLysAspAspPyrAsnArgLeuGlnGluGlnPhe-- 1219
 QY 1013 ACAGGAACCTGATGAGGCTGACACCTTGGCGGAAAGAACACCTGACACGCTGAAAGC 1072
 Db 1220 -----AspGluGlnSerLysGluAsnGluAsnIleGlyAspGlnLeuArgGln 1235
 QY 1073 TTGCTGAAAACACCTGACGAGGCTGACGCTTGGTGGAGGAGACGCTCTGCCAAAT 1132
 Db 1236 LeuGlnIleGlnValArgGluSerIleAspGlyLysLeuProSerThrAspGlnGln 1255
 QY 1133 CAATGCTGGGAAAGCAACATCTGGAAGTTTGAACAGTTCACAGCAAGAAACACCTAGG 1192
 Db 1256 SerCys-----SerSerThrProGlyLeuGluProLeuPheLysAlaThrGluGln 1273
 QY 1193 AAAATTTTGGCGCTACAAAAGAACATCTGAGAAATTTTTCATGGCCAGCAAGAAAGA 1252
 Db 1274 HisHisThrGlnProValLeuGluSerAsnLeuCysProAspTrpProSerHisSerGlu 1293
 QY 1253 TCTAGGAAGATCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1312
 Db 1294 Asp----- 1294
 QY 1313 GTAACACCTAATAAACTGAAGTTTTCGAAAAAGGACATCT---AATATGATTGCATGT 1369
 Db 1295 -----AlaSerAlaLeuGlnGlnGlyThrSerValAlaGlnIleLysAla 1309
 QY 1370 CCTACAAAAGAA---ACATCTACAAAAGCAAGTACAAATGTGGATGTGAGTCTGTAGAG 1426

Db	1310	GlnLeuLysGluIleGluAlaGluLysValGluLeuGluLeuLysValSerSerThr	--- 1328
Qy	1427	CCTATATTCACTCTTTTGCACACGACTATTGAAAAATTCACAGGTGTCACAAAGTTGAG	1486
Db	1329	-----ThrSerGlnLeuThrLysLysSer	1336
Qy	1487	GARGACTTTAAATCTGTCTACCAAGATTATCTTAAGAGTGTCTGCACAGAAATTATACGTGT	1546
Db	1337	GluGluValPheGlnLeuGlnGluGlnIleAsnLysGlnGlyLeuGlu	--- 1352
Qy	1547	TTACCTGATCTACATATCAAAAAGATATCAAAACAAATCAAAATCAAAAATAGAA	--- 1600
Db	1353	-----IleGluSerLeuLysThrValSerHisGluAlaGluValHis	1366
Qy	1600	-----	1600
Db	1367	AlaGluSerLeuGlnGlnLysLeuGluSerSerGlnLeuGlnIleAlaGlyLeuGluHis	1386
Qy	1601	-----GATCAGATGTTCCCATCAGAAATCCCAACGAGAGGAA	1636
Db	1387	LeuArgGlnLeuGlnProLysLeuAspGluLeuGlnLysLeuSerLysLysGluGlu	1406
Qy	1637	GATCAAGAAATATCTTGGGATCTCGGAGTCTCTTTGAGAGTTCTGCA	--- 1684
Db	1407	AspValSerTyrLeu	1424
Qy	1685	---AAGACTCAAGTGTGTATACCTGACTCTATGTAT	--- 1717
Db	1425	IleGlnThrGluIleGlnGlnGluAspLeuIleLysAlaLeuHisThrGlnLeuGlu	1444
Qy	1718	-----CAGAAAGTAAATGGAGATAAATAGAAAGTATAGAAAGACGT	1756
Db	1445	MetGlnAlaLysGluHisAspGluArgIleLysGlnLeuGlnValGluLeuCysGluMet	1464
Qy	1757	CCTGAGAGGCCA	--- 1792
Db	1465	LysGlnLysProGluGluIleGlyGluGluSerArgAlaLysGlnGlnIleGlnArgLys	1484
Qy	1793	ATCAAAAGACTGTTCCAAATAAAGCTTTGAATTGAAGAAATGAACAAACATTGAGA	--- 1849
Db	1485	LeuGlnAlaLeuIleSerArgLysGluAlaLeuLysGluAsnLysSerLeuGlnGlu	1504
Qy	1850	-----GCAGCTCAGATGTTCCCATCAGAATCCAAACAAAGGACGATGAAGAA	1897
Db	1505	GluLeuSerLeuAlaArgGlyThrIleGluArgLeuThrLysSerLeuAlaAspValGlu	1524
Qy	1898	AATCTTGGGATCTTGAGATCCCTGTGACAGCGTTTCACAGAGGATGTGATTATACC	1957
Db	1525	SerGlnValSerAlaGlnAsn	1539
Qy	1958	AAA---CCTACATCAAAAAGAAATTCGATACCTTAAGTGAAAATTAGAAGAGTCTCCT	2014
Db	1540	ArgLeuAlaLeuLeuGlnGluGluArgAspLysLeuIleThrGluValSerLeu	1559
Qy	2015	GTTAAAGATGGTCTTCTGAAGCCCTACCTGTGGAAGGAAAGTTCTCTTCCAAATAAAGCC	2074
Db	1560	LeuGluAsnGlnSerLeuSerSerCys	1571
Qy	2075	TTAGAAATTAAGGACAGAGAAACATTCAAAGCAGAGTCTCCTGATAAAGATGCTCTTCTG	2134
Db	1572	LeuLysLeu	1587
Qy	2135	AAGCCTACCTGTGAAGAAAGTTTCTCTCCAAATAAAGCCTTAGAATTAAGGACAGA	2194
Db	1588	LysGluIleGluSerLeuLysSerSerLysIleAlaGluSerThrGluTrpGlnGluLys	1607
Qy	2194	-----	2194
Db	1608	HisLysGluLeuGlnLysGluTyrGluIleLeuLeuGlnSerTyrGluAsnValSerAsn	1627
Qy	2195	-----GAAACACTCAAGACAGACTCTCTCTGATATGAT	2227

D	b	1628	GluAlaGluArgTleGlnHisValValGluAlaValArgGlnGlnLysGlnLeuLeuTyr	1644
Q	y	2228	GGCTTCCTGAAGCCTACTCCTCTCGAAGCAA-	2257
D	b	1648	GlyLysLeuArgSerThrGluAlaAsnLysLysLeuThrGluLysGlnLeuGlnGluAla	1667
Q	y	2258	-GTTTCTCTCTCCAAAT	2272
D	b	1668	GluGlnGluMetGluGluMetLysGluLysMetArgLysPheAlaLysSerLysGlnGln	1687
Q	y	2273	AAAGCTTTAGAATTGAAGCACACAGAAAACATTCAAAAGCAGCTCAGATGTTCCTCATCAGAA	2332
D	b	1688	LysIleLeuGluLeuGluGluGluGluAsnAspArgLeuArgAlaGluValHisProAlaGly	1707
Q	y	2333	TCCAACAAGAAGATGATGAAGAAAATCTTGCGATTTTGAGAGTTTCCTTGAGACTCTC	2392
D	b	1708	AspThrAlaLysGlu-CysMetGluThrLeu	1717
Q	y	2393	TTACAGAAATCATGTGTGTTTACCACAGCGTACACAT--CAAAAAGAAATTCGATACC	2446
D	b	1718	LeuSerSerAsnAlaSerMetLysGluGluLeuGluArgValLysMetGluTyrGluThr	1737
Q	y	2447	TTAAGTCGAAATAATTAGAAGAG--TTCCTCGATAAAGATGGTCTTCCTGAAGACCTACTGT	2503
D	b	1738	LeuSerLysLysPheGlnSerLeuMetSerGluLysAspSerLeuSerGluGluValGln	1757
Q	y	2504	GGATGAAAATTTCTCTTCCAAAT-----AAAGCCTTAGAA	2539
D	b	1758	AspLeuLysHisGlnIleGluAspAsnValSerLysGlnAlaAsnLeuGluAlaThrGlu	1777
Q	y	2540	TTGAAGGCACAGAGAAACATTCAAAAGCAGAGAT-----GTGAGT	2578
D	b	1778	LysHisAspAsnGlnThrAsnValThrGluGluGlyThrGlnSerIleProGlyGluThr	1797
Q	y	2579	TCTGTAGTCCACATTTCAGTCCTTTTTGGCAACCGACT--ACTGAAAATTTCACAGTCT	2635
D	b	1798	GluGluGlnAspSerLeuSerMetSerThrArgProThrCysSerGluSerValProSer	1817
Q	y	2636	ACAAAA-----GTTGAGAGAGACTTTAATCTTACTACCAAG---GAGGGA	2677
D	b	1818	AlaLysSerAlaAsnProAlaValSerLysAspPheSerSerHisAspGluIleAsnAsn	1837
Q	y	2678	GCACAAGACAGTAAGTGCACACAGACAGACGTCATATTGGCATTTATGGCATTTATGAACGAGCTCCA	2737
D	b	1838	TyrLeuGlnGlnIleAspGlnLeuLysGluArgIleAlaGlyLeuGluGlu-----	1854
Q	y	2738	CAAGATCAACAATAAGATGCCACATCAGAAATTAGGAAGAAAGAACATACAAA--	2794
D	b	1955	GluLysGlnLysAsnLysGluPheSerGlnThrLeuGluAsnGlnLysAsnThrLeuLeu	1874
Q	y	2795	-----TCAACTTCAGATTTCGAGATTATCTCTGTGAGTGNATACACAGAATTATGAG	2845
D	b	1875	SerGlnIleSerThrLysAspGlyGluLeuLysMetLeuGlnGluGluValThrLysMet	1894
Q	y	2846	TGTTTACTCGAGGTACATATCAAAAGAATAAAGACAACAANTGCCAAATAAGAGAG	2905
D	b	1895	AsnLeuLeuAsnGlnGlnIleGlnGluGluLeuSerArgValThr---LysLeuLysGlu	1913
Q	y	2906	TCTCCTGAAAACGCTTCCTACTTTGACGCTGCCACTGAAATGCAAAACTCTGTTCCTCAAAT	2965
D	b	1914	ThrAlaGlu-----GluGluLysAspLeuGluGluArgLeuMetAsn	1928
Q	y	2966	AAAGCCTTAGAATCGAAGAATAACAAACATCTGAGACGAGTTCAACTACCTCTATCAAAA	3025
D	b	1929	Gln-----LeuAlaGlu	1932
Q	y	3026	ATCTTGATGCATTCCTTCTTGTGAAGAGGAGGGAATTAAAAAGAAATACTGTGAA	3085
D	b	1933	LeuAsnGlySerIle-----GlyAsnTyrCysGln	1942
Q	y	3086	CAAAATTACA-----GCAAAATCGAACAATAAGTAABATAGTTTTCTGTGACTACAA	3136
D	b	1943	AspValThrAspAlaGlnIleLysAsnGluLeuLeuGluSerGluMetLysAsnLeuLys	1962

Qy	1205	CCTACAAAGAACATCTGAGAAATTTTCATGCCACAAAGAAAGATCTAGGAAGATC	1264
Db	1020	LysLeuLysAsnLysHisGluAlaMetIleThrAspLeuGluGluArgLeuArgLeuArg---	1038
Qy	1265	ACATGGCAGGAAAAAGAAACATCTGTTAAAGACTGAATCGCTGGCAGGAGTAAACACCTAAT	1324
Db	1039	-----GluGluLysGlnArgGlnGluLeuLys-----ThrArgArg	1051
Qy	1325	AAAACTGAAGTTTGGAAAAAGGAACATCTTAATATGATTGCA	1366
Db	1052	LysLeuGluGlyAspSerThrAspLeuSerAspGlnIleAlaGluLeuGlnAlaGlnIle	1071
Qy	1367	-----TGTCCTACAAAGAAACATCTACAAAGCAAGTACAAAT	1405
Db	1072	AlaGluLeuLysMetGlnLeuAlaLysGluGluLeuGlnAlaLeuAlaArg	1091
Qy	1406	GTGGATCTGAGTTCCTGAGAGCCTATATTTCAGTCTTTTGGCACACGCACTATTGAAAT	1465
Db	1092	ValGluGluGluAlaAlaGlnLysAsnMetAlaLeuLysLysIleArgGluLeuGlu---	1110
Qy	1466	TCACAGTGTACAAAAGTTGAGGAGACTTTAACTTCTGCTACCAAGATTATCTCTAAGAGT	1525
Db	1111	SerGlnIleSerGluLeuGlnGluAspLeuGlu-----SerGluArg	1124
Qy	1526	GCTGCACAGAAATTACGTGTTTACCTGATGCTACATATCAAAAGAGATACAAAACAATA	1585
Db	1125	AlaSerArgAsnLysAlaGluLysGlnLysArgAspLeuGluGluLeuGluAlaLeu	1144
Qy	1586	AATCACAAAATAGAAAGATCAGATGTTCCCATCA	1627
Db	1145	LysThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGluLeuArgSerLys	1164
Qy	1628	CGACAGCAAGATGAAGAATATCTCTGGAGTCTCTTTGAGAGTCTCTGCAAAG	1687
Db	1165	ArgGluGlnGluAlaAsnIle-----LeuLysLysThrLeuGluGluGluAlaLys	1181
Qy	1688	ACTCAAGTGTGTATACCTGAGTCTATCTATCAGAAAGCTAATGGAGATAAATAGAGAAGTA	1747
Db	1182	ThrHisGluAlaGlnIleGlnGluMetArgGlnLysHisSerGlnAlaValGluGluLeu	1201
Qy	1748	GAAGAGCTTCCTGAGAAAGCCATCTGCCTTCAGAGCTCCGTGNGAATGCAAAAG---ACT	1804
Db	1202	AlaGluGlnLeuGluGlnThrLysArgValLysAlaAsnLeuGluLysAlaLysGlnThr	1221
Qy	1805	GTTCCAAATAAAGCCTTTGAATTCAGAAATGAACAAACATTTAGAGCAGCTCAGATGTC	1864
Db	1222	LeuGluAsnGluArgGlyGluLeuAlaAsnGluValLysValLeuLeuGlnGlnGlyArg	1241
Qy	1865	CCATCAGATCCAAACAAAAGCAGATGAAGAAATTTCTGGGATTCGAGAGTCCTCGT	1924
Db	1242	AspSerGluHisLysArgLysLysValGluAlaGlnLeuGlnGluGlnValLysPhe	1261
Qy	1925	---CAGACGGTTTCACAGAAGGATGTGTAATTTACCCAAAGCTACACAT---CAAAAGAA	1978
Db	1262	AsnGluGlyGluArgValArgThrGluLeuAlaAspLysValThrLysLeuGlnValGlu	1281
Qy	1979	TTCGATACCTTAAGTGGAAATTAGAAGAGTCTCCTCTTAAGATGCTCTTCGAGCCCT	2038
Db	1282	LeuAspAsnValThrGlyLeuLeuSerGlnSerAspSerLysSerLysLeu-----	1299
Qy	2039	ACCTGTGGAGGAAGTTTCTCTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACA	2098
Db	1300	-----ThrLysAspPheSerAla	1305
Qy	2099	TTCAAAGCAGAGTCTCTGTGATAAGATGGTCTCTCTGAAGCCTACCTCTGGAAGGAAGTT	2158
Db	1306	LeuGluSerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluAsnArgGlnLysLeu	1325
Qy	2159	TCTCTTCCAAATAAAGCCTTAGAATTAAGGAC-----ACAGAAACACTC	2203
Db	1326	SerLeuSerThrLysLeuLysGlnValGluAspGluLysAsnSerPheArgGluGlnLeu	1345

QY	2204	AAACGAGGTCTCTGGTGAATGATGGTCCTCTCGAAG-----CCTACTCTGTGAAGG	22525
Db	1346	GluGluGluGluGluGluAlaLysHisAsnLeuLeuLysGlnIleAlaThrLeuHisAla	1365
QY	2255	AAAGTTCTCTCTCCAAAATAAAGCTTTAGAA-----TTGAAGAGCAGAGAA	2299
Db	1366	GlnValAlaAspMetLysLysLysMetGluAspSerValGlyCysLeuGluThrAlaGlu	1385
QY	2300	ACATTCAAGCAGCTCAGATGTTCCCATCAGAA-----TCCAACAAGAGGATGATGAA	2353
Db	1386	GluValLysArgLysLeuGlnLysAspLeuGluGlyLeuSerGlnArgHisGluGluLys	1405
QY	2354	GAATAATCTCTGGGATTTTGAGAGTTTCCTTCAGACTCTCTTACAGAAT-----GATGTG	2407
Db	1406	ValAlaAlaTyzAspLysLeuGluLysThrLysThrArgLeuGlnGluLeuAspAsp	1425
QY	2408	TGTTTTACCAGGCCTACACAT-----CAAAAA	2434
Db	1426	LeuLeuValAspLeuAspHisGlnArgGlnSerAlaCysAsnLeuGluLysLysGlnLys	1445
QY	2435	GAATTCGAT-----ACCTTAAGTCGAAATTTAGAAGAGTCTCCT	2473
Db	1446	LysPheAspGlnLeuLeuAlaGluGluLysThrIleSerAlaLysTyraAlaGluLys	1465
QY	2474	GATAAGATGGTCTCTTCAAGCCCTACCTGTGGAATGAAATTTCTCTCCAAATAAAGCC	2533
Db	1466	AspArg-----AlaGlu	1469
QY	2534	TTAGAATTGAAGNACAGAGAACATTTCAAAGCAGGATGTGATCTCTGTAGAGTCCACA	2593
Db	1470	AlaGluAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArgAlaLeuGluGluAla	1489
QY	2594	TTCACTCTTTTGGCAACACCGACTACTCGAAATTCACAGCTCTACAAAGTTGAGGAAGAC	2653
Db	1490	MetGluGlnLysAlaGluLeuGluArgLeuAsnLysGlnPheArgThrGluMetGluAsp	1509
QY	2654	TTTAATCTTACTACAGAGGAGGAGAACAAAGACAGCTAATCTGGCAACAGGAACGTGAT	2713
Db	1510	Leu---MetSerSerLysAspValGlyLysSerVal-----HisGlu	1523
QY	2714	ATTGGCATTTTGAACGAGCTCCACAAGATCAACAATTAAGATGCCACATCAGAAATTA	2773
Db	1524	LeuGluLysSerLysAlaGalaLeuGluGlnGlnValGluGluMetLysThrGlnLeuGlu	1543
QY	2774	GGAGAAAAAGAGATACAAAATCAAACTTCAGATTTCTGAGATT---ATCTCTGTGAGTGT	2830
Db	1544	GluLeuGluAspGluLeuGlnAlaThrGluAspAlaLysLeuArgLeuGluVal-----	1561
QY	2831	ACACAGAATTATGAGTGTTTTACCTGAGGCTCATATCAAAAGNAATTAAGACACAAAT	2890
Db	1562	----AsnLeuGlnAlaMet---LysAlaGlnPheGluArgAspLeuGln-----	1575
QY	2891	GGCAAAATAGAAGGTCTCTCTGAAAAGCCTTCTCACATT-----GAGCCT	2935
Db	1576	GlyArgaspGluGlnSerGluGluLysLysGlnLeuValArgGlnValArgGluMet	1595
QY	2936	GCCACTGAATGCAAACTCTGTTTCCAAATAAAGGCTTAGAATGGAAGAAATAACAACA	2995
Db	1596	GluAlaGluLeuGluAspGluArgLysGlnArgSerMetAlaValAlaAlaArgLysLys	1615
QY	2996	TTGAGACGAGATTCAACTACCCCTATCAAAATCTTGGATGCA-----	3037
Db	1616	LeuGluMetAspLeuLysAspLeuGluAlaHisIleAspSerAlaAsnLysAsnArgAsp	1635
QY	3038	-----CTTCCTCTCTTGTGAAGAGGAAGG	3061
Db	1636	GluAlaIleLysGlnLeuArgLysLeuGlnAlaGlnMetLysAspCysMetArgGluLeu	1655
QY	3062	GAACCTTAAAAAGATACTGTGAACAATAATCAGCA-----AAA	3100
Db	1656	AspAspThrArgAlaSerArgGluIleuAlaGlnAlaLysGluAsnGluLysLys	1675
QY	3101	ATGAAACAAATGAAAAATAAGTTTGTGTACTACAAAAGCAACTGTCTCAGAAGCGAAGAA	3160

```

1676 LeuLysSerMetGluAlaGluMetIleGlnLeuGlnGluGluLeuAlaAlaGluArg 1695
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
3161 ATAAATACAGTGTAGAGAACCAAAAGCTAAATGGGAACAAGAGCTCTGCAGTGTG--- 3217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1696 AlalysArgGlnAlaGlnGlnGluArgAspGluLeuAlaAspGluIleAlaSerSer 1715
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3218 -----AGATTGCTTTAAATCAAGAAGAGAGAGAGAGAAATGTCGATATATTA 3268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1716 GlyLysGlyAlaLeuAlaLeu-----GluGluLysArg 1728
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3269 AAAGAAAAATTT---AGACCCGAGAGCAACTTAGGAAAAAGTTAGAAAGTGAACACCAA 3325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1729 GluAlaArgIleAlaGlnLeuGluGluGluGluGluGluGluGluGlnGlnGlnLeu 1748
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3326 CTTGAACAGACTCTCAGATATCAAGATATAGAAATGAAAGTGTACAAAGTAAATTTGAAT 3385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1749 IleAsnAspArgLeuLysLysAlaAsnLeuGlnIleAspGlnIleAsnAlaAspLeuAsn 1768
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3386 ---CAGTTTCTCACACTCATGAAGTGAATGATCTCTTTTCATGAAAATTCATGTTG 3442
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1769 LeuGluArgGlyHisAlaGlnLysAsnGluAsn-----AlaArgGlnGlnLeu 1784
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3443 AAAAGAAAAATTTGCACTGCTAAAGCTGGAAGTAGCCACACTGAAACATCAACACCAAGGTG 3502
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1785 GluArgGlnAsnLysGluLeuLysValLys-----LeuGlnGluMetGluGlyThr 1801
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3503 AAGAAAAATTAATCTTTGAGGACATTAAGATTTTACAGAAAGAAATGCTGAACCTCAA 3562
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1802 ValLysSerLysTyrLysAlaSerIleThrAlaLeuGluAlaLysIleAlaGlnLeuGlu 1821
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3563 ATGACCCCTA----- 3571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1822 GluGlnLeuAspAsnGluThrLysGluArgGlnAlaAlaCysLysGlnValArgArgThr 1841
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3572 -----AAACTGAAACAGAAAAACAGTA-----ACAAAAAGGCATCTCAG 3610
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1842 GluLysLysLysLysAspValLeuLeuGlnValAspAspGluArgAsnAlaGluGln 1861
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3611 TATAGACAGCAGCTTAAAGTCTGCGGCAGACACGATGCTGACTCTTAAATTTGAAG 3670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1862 TyrLysAspGlnAlaAspLysAlaSerThrArgLeuLysGlnLeuLysArgGlnLeuGlu 1881
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3671 GAA 3673
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1882 Glu 1882

RESULT 14
152300
gi|152300|giantin - human
N;Alternate names: gcp372
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C;Accession: I52300
R;Shoda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A;Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A;Reference number: I52300; MUID:95100974; PMID:7802676
A;Accession: I52300
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3225 <RES>
A;Cross-references: GB:D25542; NID:9662389; PIDN:BA05025.1; PID:9808869
C;Superfamily: giantin

Alignment Scores:
Pred. No.: 9.95e-10 Length: 3225
Score: 310.00 Matches: 235
Percent Similarity: 36.62% Conservative: 211
Best Local Similarity: 19.29% Mismatches: 426
Query Match: 4.61% Indels: 346
DB: 2 Gaps: 48
```

```

US-09-602-362E-26 (1-3673) x I52300 (1-3225)
QY 680 GTGCAAAAACAGGCTAGCTCACCCTTACTGCGCCATACAGAAAAGAACGACAA 739
      ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
1067 LeuGlnAspLysThrAsnGlnIleAspLeuLeuGlnAlaGluIleSerGluAsnGlnAla 1086
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
740 ACTGTGGATTTTACTTAACCAAAAATGCAAAATGCAAAATGCAAAATGCAAAATGCA 790
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1087 IleIleGlnLysLeuIleThrSerAsnThrAspAlaSerAspGlyAspSerValAlaLeu 1106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
791 TCTAAATGCACAGCCCTCATCTTCCCATATGTGAAGCTCATCAGAGATAGTCGGCATG 850
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1107 ValLysGluThrValValIleSerProCysThrGlySerSerGluHisTrpLysPro 1126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
851 CTTCTTACGCAAAATGTT-----GACGCTTTTCTGAAGCATACATGGAATAACT 901
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1127 GluLeuGluGluLysIleLeuAlaLeuGluLysGlnLysGluGlnLeuGlnLysLysLeu 1146
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
902 GCAGAACGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 GlnGluAlaLeuThrSerArgLysAlaIleLeuLysLysAlaGlnGlnLysGluArgHis 1166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
962 ATACGA-----AAATTACCTAAATACTCTCAAAATCCATCCAGAAAGAACATCT 1012
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1167 LeuArgGluGluLeuLysGlnLysAspTyrAsnArgLeuGlnGlnGlnPhe--- 1185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1013 ACAGGAACCTGTGATGAGGCTGCACCTTGGCGGAAAGAACACCTGACACGGCTGAAAGC 1072
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1186 -----AspGluGlnSerLysGluAsnGluAsnIleGlyAspGlnLeuArgGln 1201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1073 TTGCTGGAAAAACACCTGCAGCGCTGCACGCTGTGTGTGGAGGAACTGTGCCAAAT 1132
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1202 LeuGlnIleGlnValArgGluSerIleAspGlyLysLeuProSerThrAspGlnGlnGlu 1221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1133 CAATGCTCTGGGGAAGAACATCTGGAAGTTTGAACAGTCAACAGAGAAACACCTAGG 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1222 SerCys-----SerSerThrProGlyLeuGluGluProLeuPheLysAlaThrGlnGln 1239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1193 AAAATTTTGGGCTTACAAAAGAACATCTGAGAAATTTTCATGGCCAGCAAGAAAGA 1252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1240 HisHisThrGlnProValLeuGluSerAsnLeuCysProAspTrpProSerHisSerGlu 1259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1253 TCTAGGAAGATCACATGGGAGGAAAGAAACATCTCTAAAGACTGAATCGCTGGCAGGA 1312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1260 Asp----- 1260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1313 GTAAACACCTAATAAACTGAAAGTTTGGAAAAAGGAACATCT---AATATGATTGTCATG 1369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1261 -----AlaSerAlaLeuGlnGlnGlyThrSerValAlaGlnIleLysAla 1275
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1370 CCTACAAAAGAA---ACATCTTACAAAAGCAAGTCAAAATGTGGATGTGAGTCTGTAGAG 1426
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1276 GlnLeuLysGluIleGluAlaGluLysValGluLeuLysValSerSerThr--- 1294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1427 CCTATTTCAGTCTTTTGGCACCGGACTATTGAAAATTCACAGTGTACAAAAGTTGAG 1486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1295 -----ThrSerGluLeuThrLysLysSer 1302
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1487 GAAGACTTTAATCTTGTACCAAGATTATCTCTAAGAGTCTGTGCACAGAAATATACGTGT 1546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1303 GluGluValPheGlnLeuGlnGlnIleAsnLysGlnGlyLeuGlu----- 1318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1547 TTACTGTGCTACATATCAAAAAGATATCAAAACATTAATCAAAAATAGAA----- 1600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1319 -----IleGluSerLeuLysThrValSerHisGluAlaGluValHis 1332
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1600 ----- 1600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1333 AlaGluSerLeuGlnGlnLysLeuGluSerSerGlnLeuGlnIleAlaGlyLeuGluHis 1352
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1601 -----GATCAGATGTTCCCATCAGATATCAAAACAGAGGAA 1636
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1353 LeuArgGluLeuGlnProLysLeuAspGluLeuGlnLysLeuIleSerLysLysGluGlu 1372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Qy	1637	GATGAGAAATATCTCTGGGATTCCTGGAGTCTCTTTGAGAGTTCGTCA	-----1684
Db	1373	AspValSerTyrLeu-----SerGlyGlnLeuSerGluLysGluAlaLeuThrLys	1390
Qy	1685	--AAGACTCAAGTGTATACCTCAGTCTATGTAT	-----1717
Db	1391	IleGlnThrGluIleIleGluGlnGluAspLeuIleLysAlaLeuHisThrGlnLeuGlu	1410
Qy	1718	-----CAGAAAGTAATCGAGATAAATAGAGAAGTAGAAGAGCTT	1756
Db	1411	MetGlnAlaLysGluHisAspGluArgIleLysGlnLeuGlnValGluLeuCysGluMet	1430
Qy	1757	CCTGAGAGCCA-----TCGCTTCAAGCTGCGCTNGAA	-----1792
Db	1431	LysGlnLysProGluGluIleGlyGluGluSerArgAlaLysGlnIleGlnArgLys	1450
Qy	1793	ATGCAAAAGACTGTTCCAAATAAAGCCTTTCAATTGAAGATCAACAAACATTTCAGA	-----1849
Db	1451	LeuGlnAlaLeuIleSerArgLysGluAlaLeuLysGluAsnLysSerLeuGlnGlu	1470
Qy	1850	-----GCAGTCAGATGTTCCCATCAGAAATCCAAACAAAGACGATGAAGAA	1897
Db	1471	GluLeuSerLeuAlaArgGlyThrIleGluArgLeuThrLysSerLeuAlaAspValGlu	1490
Qy	1898	AATTCTTGGGATCTCGAGAGTCCCTGTGAGACGGTTTCACAGAGGATGTATTACCC	1957
Db	1491	SerGlnValSerAlaGlnAsn-----LysGluLysAspThrValLeuGly	1505
Qy	1958	AAA---CCTACATCACTAAAAAGAAATTCGATACCTTAAGTGGAAATAGAGAGTCTCCT	2014
Db	1506	ArgLeuAlaLeuLeuGlnGluArgAspLysLeuIleThrGluMetAspArgSerLeu	1525
Qy	2015	GTTAAAGATGGTCTTCTGAAGCCTACTCTGTGGAAGAAAGTTCTCTCCCAATAAGACC	2074
Db	1526	LeuGlnAsnGlnSerLeuSerSerCys-----GluSer	1537
Qy	2075	TTGAATTAAGGACAGAGAAACATTCAAAGCAGAGTCTCCTGATGAAGATGCTTCTG	2134
Db	1538	LeuLysLeu-----AlaLeuGluGlyLeuThrGluAspLysGluLysLeuVal	1553
Qy	2135	AGCCTACTGTGGAAGAAAGTTTCTCTCCAAATAAAGCCTTAGAATTAAGGACAG	2194
Db	1554	LysIleLeuGluSerLeuLysSerSerLysIleAlaGluSerThrGluTrpGlnGluLys	1573
Qy	2194	-----	-----2194
Db	1574	HisLysGluLeuGlnLysGluTyrGluIleLeuLeuGlnSerTyrGluAsnValSerAsn	1593
Qy	2195	-----GAAACACTCAAGCAGAGTCTCCTCGATAATGAT	2227
Db	1594	GluAlaGluArgIleGlnHisValValGluAlaValArgGlnGluLysGlnGluLeuTyr	1613
Qy	2228	GGTCTTCTGAAGCCTACTCTGGAAGGAAA-----	-----2257
Db	1614	GlyLysLeuArgSerThrGluAlaAsnLysLysGluThrGluLysGlnLeuGlnGluAla	1633
Qy	2258	-----GTTTCTCTTCCAAAT	2272
Db	1634	GluGlnGluMetGluGluMetLysGluLysMetArgLysPheAlaLysSerLysGlnGln	1653
Qy	2273	AAAGCTTTAGAAATTAAGGACAGAGAAACATTCAAGCAGCTCAGATGTTCCCATCAGAA	2332
Db	1654	LysIleLeuGluLeuGluGluGluAsnAspArgLeuArgAlaGluValHisProAlaGly	1673
Qy	2333	TCCAAACAAAGGATGATGAAGAAAATTCCTGGCATTTTCAGAGATTTCCCTTGAGACTCTC	2392
Db	1674	AspThrAlaLysGlu-----CysMetGluThrLeu	1683
Qy	2393	TTACAGAAATGATGTGTTTACCAAGGCTACAT-----CAAAAGAAATTCGATACC	2446
Db	1684	LeuSerSerAsnAlaSerMetLysGluGluLeuGluArgValLysMetGluTyrGluThr	1703

Qy	2447	TTAAGTGGAAAAATTAGAAGAG---TCCTCGATFAAAGATGGTCTTCTTGAAGCCTACCTGT	2503
Db	1704	LeuSerLysLysPheGlnSerLeuMetSerGluLysAspSerLeuSerGluGluValGln	1723
Qy	2504	GGAATGAAATTTCTCTCCAAAT-----AAAGCCCTTAGAA	2539
Db	1724	AspLeuLysHisGlnIleGluGlyAsnValSerLysGlnAlaAsnLeuGluAlaThrGlu	1743
Qy	2540	TTGAAGCAGACAGAAAACATTCAAAGCAGAGAT-----GTGAGT	2578
Db	1744	LysHisAspAsnGlnThrAsnValThrGluGluGlyThrGlnSerIleProGlyGluThr	1763
Qy	2579	TCTGTAGAGTCCACATTTCAGTCTTTTGGCAACCGACT--ACTGAAAAATTCAGACT	2635
Db	1764	GluGlnAspSerLeuSerMetSerThrArgProThrCysSerGluSerValProSer	1783
Qy	2636	ACAAAA-----GTTGAGAGAGACTTTTAATCTTACTACCAAG--GAGGGA	2677
Db	1784	AlaLysSerAlaAsnProAlaValSerLysAspPheSerSerHisAspGluIleAsnAsn	1803
Qy	2678	GCAACAAGACAGTAACGTGCACACAGACGACGTATTTGGCATTTATTGAACGAGCTCCA	2737
Db	1804	TyrLeuGlnGlnIleAspGlnLeuLysGluArgIleAlaGlyLeuGluGlu-----	1820
Qy	2738	CAAGATCAAAACAATAAGATGCCACACATCAGAATTAGGAAGAAAGAGATACAAAA	2794
Db	1821	GluLysGlnLysAsnLysGluPheSerGlnThrLeuGluAsnGluLysAsnThrLeuLeu	1840
Qy	2795	-----TCAACTTCAGATTCTGAGATTATCTCTGTGAGTGATACACAGAATTATGAG	2845
Db	1841	SerGlnIleSerThrLysAspGlyGluLeuLysMetLeuGlnGluGluValThrLysMet	1860
Qy	2846	TGTTTACTGAGGCTACATATCAAAAGAATAAAGACACAAATGCGAAAATAGAGAG	2905
Db	1861	AsnLeuLeuAsnGlnGlnIleGlnGluGluLeuSerArgValThr--LysLeuLysGlu	1879
Qy	2906	TCTCCTGAAAAGCCCTCTCACTTTGAGCGCTGCCACTGAAATCGAAAATCTGTTCCTCAAT	2965
Db	1880	ThrAlaGlu-----GluGluLysAspLeuGluGluArgLeuMetAsn	1894
Qy	2966	AAAGGCTTAGAATCGAAGATAAACACACATTTGAGACGAGATTCAACTCCCTATCAAAA	3025
Db	1895	Gln-----LeuAlaGlu	1898
Qy	3026	ATCTTGGATGCATCTCTTCTTGTGAAGAGAGGGAACCTTAAAAAGATAACTGTGCAA	3085
Db	1999	LeuAsnGlySerIle-----GlyAsnTyrCysGln	1908
Qy	3086	CAAAATTACA-----GCMAAATCGAACAAATGAAAATAAGTTTGTGTACTACAA	3136
Db	1909	AspValThrAspAlaGlnIleLysAsnGluLeuLeuGluSerGluMetLysAsnLeuLys	1928
Qy	3137	AAGCAACTGTGCAAGCGAAAGAATAAAATTCACAGTTAGAGAACCAAAAAGCTTAAATGG	3196
Db	1929	LysCysValSerGluLeuGluGluGlyGlnLeuValLysGluLysThrLysVal	1948
Qy	3197	GAACAAGAGTCTCGACTGTGAGATTG-----CCTTTA	3229
Db	1949	GluSerGluIleArgLysGluTyrLeuGluLysIleGlnGlyAlaGlnLysGluProGly	1968
Qy	3230	AATCAAGAAGACAGAGAGAAATGTCGATATATTAAAGAAAAAATTAGACCCGAA	3289
Db	1969	AsnLysSerHisAlaLysGluLeuGln--GluLeuLeuLysGluLysGlnGlnGluVal	1987
Qy	3290	GAGCAACTTAGGAAAAAGTTAGAAAGTGAAACACCAA-----CTTTGAACAG	3334
Db	1988	LysGlnLeuGlnLysAspCys--IleArgTyrGlnGlnLysIleSerAlaLeuGluArg	2006
Qy	3335	ACTCTCAGAATACAGATATA-----GNAATTGAAAAGTGTAACA	3373
Db	2007	ThrValLysAlaLeuGluPheValGlnThrGluSerGlnLysAspLeuGluIleThrLys	2026
Qy	3374	AGTAATTTGAATCAG--GTTTCTCACACATCAAGAGTGAATGATCTC-----	3421

Db 2027 GluAsnLeuAlaGlnAlaValGluHisArgLysLysAlaGlnAlaGluLeuAlaSerPhe 2046
QY 3422 -----TTTCATGAAATTCATG 3439
Db 2047 LysValLeuLeuAspThrGlnSerGluAlaAlaArgValLeuAlaAspLeuLys 2066
QY 3440 TTGAAAGGAAATTT-----GCCATGCTAAACTGGAAGTA 3475
Db 2067 LeuLysLysGluLeuGlnSerAsnLysGluSerValLysSerGlnMetLysGlnLysAsp 2086
QY 3476 GCCACACTGAACATCAACACAGGTGACGAGAAATAAATACCTTCAGGACATTAAGATT 3535
Db 2087 GluAspLeuGluArgLeuGluGlnAlaGluGluHisLeuLysGluLysLysAsn 2106
QY 3536 TTACAGAAAG-----AATGCT 3553
Db 2107 MetGlnGluLysLeuAspAlaLeuArgArgGluLysValHisLeuGluGluThrIleGly 2126
QY 3554 GAACCTCAATGACCTTAAACTGAAACAGAAACAGTAACAAAAGGCATCTCAGTAT 3613
Db 2127 GluIleGlnValThrLeuAsnLysLysAspLysGluVal-----GlnGlnLeu 2142
QY 3614 AGACAGCAGCTTAAAGTCTCAGCGCAGACACAGCATGCTGACTTCTAAATTG 3667
Db 2143 GlnGluAsnLeuAspSerThrValThrGlnLeuAlaAlaPheThrLysSerMet 2160

RESULT 15
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A:Reference number: JC5837; PMID:98033490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <OK>
A:Cross-references: DDBJ:D25543; NID:9516825; PIDN:BA05026.1; PID:9516826
C:Comment: This protein plays a role in the formation and maintenance of the characteris
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Alignment Scores:
Pred. No.: 2.59e-09 Length: 3187
Score: 302.50 Matches: 253
Percent Similarity: 37.56% Conservative: 197
Best Local Similarity: 21.12% Mismatches: 451
Query Match: 4.69% Indels: 297
DB: 2 Gaps: 54

US-09-602-362E-26 (1-3673) x JC5837 (1-3187)

QY 641 ATGTGGCAACACTGCTGCTATGTGTCAGTCATCAGGTGCAAAACAG----- 691
Db 1453 LeuThrLysSerLeuAlaAspValGluSerGlnValSerValGlnAsnGlnGluLysAsp 1472
QY 692 GCTAGCTCACCCCTTTACTGGCCATACAGAAAGACGACCAACTGTGGAATTT 751
Db 1473 AlaLeuLeuGlyLysLeuAlaLeuLeuGlnGluGluArgAspLysLeuLeuValGlu--- 1491
QY 752 TTACTACAAATAATGCAAAATGCAATTAATAGTCTAAATGACGACCCCTCATG 811
Db 1492 ---MetAspLysSerLeuLeuGluAsnGlnSerLeuGlyGlySerCysGluSerLeuLys 1510
QY 812 CTTGCCATATGTGAAGCTCATCAGATAGTGGCATGCTTCTTTCAGCAAAATGTGAC 871
Db 1511 LeuAlaLeu-----GlyGly-----LeuThrGluAspLysGlu 1521

QY 872 GTCTTTGCTGAGACATACATGGAATAACTGCAGAACGTTATGCTGCTCGTGGAGTT 931
Db 1522 LysLeuMetLysGluLeuLeuSerValArgCysSerLysIleAlaGluSerThrGluTrp 1541
QY 932 AATTACATTTCATCAACAACCTTTTGGAAACATATACGAAATTTACCTAAAAATTCCTCAAAAT 991
Db 1542 GlnGluLysHisLysGluLeuGlnLysLysLysLysLysLysLysLysLysLysLysLys 1561
QY 992 ACCAATCCAGAAAGAACATCTACAGGAACACCTGATGAGGCTGCACCTTGGCGGAAGA 1051
Db 1562 ValSerAsnGlu-----AlaGluArg 1568
QY 1052 ACACCTGACACGGCTGAAAGCTTCTGCAAAAAACACCTGACGAGGCTGCACGCTTGGT 1111
Db 1569 IleGlnHisValValGluSerValArgGlnGluLysGlnGluValTyAlaLysLeuArg 1588
QY 1112 GAGGAAAGCTCTGCCAAATTCATGTCGGGAAACACATCTGGAAGTTTGAACAG 1171
Db 1589 SerAlaGluSerAspLysArgGluArgGluLysGlnLeuGlnAspAla-----GluGln 1606
QY 1172 TCAACAGAGAAACACCTAGGAAATTTTGGAGGCTTACAAAAGAACATCTGAGAAAATTT 1231
Db 1607 GluMetGluGluMetLysGluLysMetArgLysPheAlaLysSerLysGlnGlnLysIle 1626
QY 1231 ----- 1231
Db 1627 LeuGluLeuGluGluGluAsnAspArgLeuArgAlaGluAlaGlnProValGlyGlyAla 1646
QY 1232 -----TCATGGCCAGCAAAAGAAAGATCTAGGAAG 1261
Db 1647 AsnGluSerMetGluAlaLeuLeuSerSerAsnAlaSerLeuLysGluGluLeuGluArg 1666
QY 1262 ATCATCGGAGGAAAGAAACATCTGTAAGACTGAATCGCTGGCAGGAGTAACACCT 1321
Db 1667 IleThrLeuGluTyLysThrLeuSerLysGluPheGluAlaLeuMetAla----- 1683
QY 1322 AATAAAGTGAAGTTTGGAAAAAGAACATCTAATATG----- 1360
Db 1684 ---GluLysAsnThrLeuSerGluGluThrArgAsnLeuLysLeuGlnValGluAlaGln 1702
QY 1361 ---ATTGTCATCTCTACAAAAGAACATCTACAAAAGCAAGTACAAATGTGGATGTGAGT 1417
Db 1703 GluLeuLysGlnAlaSerLeuGluThrGluLysSerAspGluProLysAspVal--- 1721
QY 1418 TCTGTAGAGCCTATATTTCAGTCTTTTGGCACACGACTATTGAAAT-----TCA 1468
Db 1722 --IleGluGluValThrGluAlaValValGlyLysSerGlnGluGlnAspSerLeuSer 1740
QY 1469 CAGTGTACAAAAGTTGAGGA---GACTTTAATCTTGTCTACCAAGATTATCTCTAAGAGT 1525
Db 1741 GluAsnAlaLysLeuGluAspAlaGluAlaThrLeuLeuAlaAsnSerAlaLysProGly 1760
QY 1526 GCTGCACAGAATTATACGTGTTTACCTGATGCT---ACATATCAAAAAGATATCAAAACA 1582
Db 1761 ValSerGluThrPheSerSerHisAspAspIleAsnAsnTyLeuGlnGlnLeuAspGln 1780
QY 1583 ATAATACAAAATAGAGATCAGATGTTCCATCAGAAATCCAAACAGAGAGAGAGAGAA 1642
Db 1781 LeuLysGlyArgIleAlaGlu-----LeuGluMetGluLysGlnLysAspArg 1796
QY 1643 GAATATTCTTGG-----GATTCTGGGAGTCTCTTTGAGAGTTCTGCAAG 1697
Db 1797 GluLeuSerGlnThrLeuGluAsnGluLysAsnAlaLeuLeuThrGlnIleSerAlaLys 1816
QY 1688 ---ACTCAAGTGTATACCTGAGTCTATGATCAGAAAGTAATGAGAGATAATAGAGAA 1744
Db 1817 AspSerGluLysLeuLeuGluGluGluValAlaLysIleAsnMetLeuAsnGlnGln 1836
QY 1745 GTAGAAGAGCTTCTCGAAGACCCATCTGCCCTTCAAGCTGCGCGTNGAA----- 1792
Db 1837 IleGlnGluGluLeuSerArgValThrLysLeuLysGluThrAlaGluGluLysAsp 1856
QY 1793 ---ATGCAAAAGACTGTTTCCAAATAAAGCCTTGAATTG----- 1828

Qy 3623 -----CTTAAAGTTCTGACGGCAGAGAACACGATGCTGACTTCTAAATTGAAG 3670
Db 2524 MetSerLeuaspAlaLeuGlnGluAenGlnGlyLeuSerLysGluIleLys 2541

Search completed: July 15, 2004, 09:51:56
Job time : 233.685 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:50:37 ; Search time 171.903 Seconds
(without alignments)

13956.780 Million cell updates/sec

Title: US-09-602-362E-26

Perfect score: 6448
Sequence: 1 caagagcttgagacacaga.....tgactcttaaatgaagaa 3673

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 2570690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q=/cgn2_6/prodata/1/pubpaa/US09602362/runat_15072004_093627_22071/app_query.fasta_1.10325
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USRF=US09602362@cgn 1 1 719 @runat_15072004_093627_22071
-NCFU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOF=10 -XGAPEXT=0.5
-XGAPOF=6 -XGAPEXT=7 -XGAPOF=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

1	5938	92.1	1225	14	US-10-177-293-332	Sequence 332, App
2	5173	80.2	1011	16	US-10-408-765A-1557	Sequence 1557, App
3	5173	80.2	1239	13	US-10-007-805-577	Sequence 577, App
4	5173	80.2	1239	14	US-10-076-622-577	Sequence 577, App
5	5173	80.2	1239	14	US-10-124-805-577	Sequence 577, App
6	3458	53.6	1341	12	US-10-058-270A-4	Sequence 4, Appli
7	3458	53.6	1341	13	US-10-007-805-565	Sequence 565, App
8	3458	53.6	1341	14	US-10-076-622-565	Sequence 565, App
9	3458	53.6	1341	14	US-10-177-293-334	Sequence 334, App
10	3458	53.6	1341	14	US-10-124-805-565	Sequence 565, App
11	3440	53.3	1349	13	US-10-007-805-573	Sequence 573, App
12	3440	53.3	1349	14	US-10-076-622-573	Sequence 573, App
13	3440	53.3	1349	14	US-10-124-805-573	Sequence 573, App
14	2374	36.8	1013	13	US-10-007-805-553	Sequence 553, App
15	2374	36.8	1013	14	US-10-076-622-553	Sequence 553, App
16	2374	36.8	1013	14	US-10-124-805-553	Sequence 553, App
17	2372.5	36.8	1095	13	US-10-007-805-493	Sequence 493, App
18	2372.5	36.8	1095	14	US-10-076-622-493	Sequence 493, App
19	2372.5	36.8	1095	14	US-10-124-805-493	Sequence 493, App
20	2367	36.7	1002	9	US-09-604-287A-475	Sequence 475, App
21	2367	36.7	1002	9	US-09-551-621-475	Sequence 475, App
22	2367	36.7	1002	13	US-10-007-805-475	Sequence 475, App
23	2367	36.7	1002	14	US-10-076-622-475	Sequence 475, App
24	2367	36.7	1002	14	US-10-124-805-475	Sequence 475, App
25	1776.5	27.6	661	13	US-10-007-805-552	Sequence 552, App
26	1776.5	27.6	661	14	US-10-076-622-552	Sequence 552, App
27	1776.5	27.6	661	14	US-10-124-805-552	Sequence 552, App
28	1775	27.5	743	13	US-10-007-805-494	Sequence 494, App
29	1775	27.5	743	14	US-10-076-622-494	Sequence 494, App
30	1775	27.5	743	14	US-10-124-805-494	Sequence 494, App
31	1769.5	27.4	650	9	US-09-825-301-25	Sequence 25, Appl
32	1769.5	27.4	650	9	US-09-604-287A-469	Sequence 469, App
33	1769.5	27.4	650	10	US-09-551-621-469	Sequence 469, App
34	1769.5	27.4	650	13	US-10-007-805-469	Sequence 469, App
35	1769.5	27.4	650	14	US-10-076-622-469	Sequence 469, App
36	1769.5	27.4	650	14	US-10-124-805-469	Sequence 469, App
37	1769.5	27.4	650	14	US-10-033-527-25	Sequence 25, Appl
38	1542	23.9	512	16	US-10-181-663-16	Sequence 16, Appl
39	1414.5	21.9	445	9	US-09-825-301-23	Sequence 29, Appl
40	1414.5	21.9	445	9	US-09-604-287A-473	Sequence 473, App
41	1414.5	21.9	445	10	US-09-551-621-473	Sequence 473, App
42	1414.5	21.9	445	13	US-10-007-805-473	Sequence 473, App
43	1414.5	21.9	445	14	US-10-076-622-473	Sequence 473, App
44	1414.5	21.9	445	14	US-10-124-805-473	Sequence 473, App
45	1414.5	21.9	445	14	US-10-033-527-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-10-177-293-332
; Sequence 332, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavaru, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,

; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER

; FILE REFERENCE: MRI-038

; CURRENT APPLICATION NUMBER: US/10/177,293

; CURRENT FILING DATE: 2002-06-21

; PRIOR APPLICATION NUMBER: US 60/299,887

; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 60/301,572

; PRIOR FILING DATE: 2001-06-27

; PRIOR APPLICATION NUMBER: US 60/306,501

; PRIOR FILING DATE: 2001-07-18

; PRIOR APPLICATION NUMBER: US 60/325,002

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/362,585

; PRIOR FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: US 60/xxx,xxx

; PRIOR FILING DATE: 2002-05-14

; NUMBER OF SEQ ID NOS: 506

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 332

; LENGTH: 1225

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-177-293-332

Alignment Scores:

Pred. No.: 0 Length: 1225
Score: 5938.00 Matches: 1159
Percent Similarity: 99.57% Conservative: 1
Best Local Similarity: 99.48% Mismatches: 4
Query Match: 92.09% Indels: 1
DB: 14 Gaps: 0

US-09-602-362E-26 (1-3673) x US-10-177-293-332 (1-1225)

QY 181 ATGAAGAGGCTCTAGTGGCTGGCAAGGGCGTGGGGGGCCGGAGCCCGCAACCCC 240
Db 1 MetLysArgLeuLeuAlaAlaGlyLysGlyValArgGlyProGluProAsnPro 20
QY 241 TTGAGGAAGCGGTCTACACTGAGAGGACTAGCGGACCCTACTTCGGGATCTAGG 300
Db 21 PheSerGluArgValThrGluLysAspTyrGlyThrIleTyrPheGlyAspLeuGly 40
QY 301 AAGATCCATACAGTGGCTCCCGGGCCAAAGTCCAGAAGCTGGAGAAGATGACAGTAGG 360
Db 41 LysIleHisThrAlaAlaSerArgGlyGlnValGlnLysLeuGluLysMetThrValGly 60
QY 361 AAGAAGCCCTCAACTGGAACAAAGAGATATGAAGAAGAGACTGCTCTACACTGGGCC 420
Db 61 LysLysProValAsnLeuAsnLysArgAspMetLysLysArgThrAlaLeuHisTrpAla 80
QY 421 TGTGTCAATGCCATGCANAGTAGTAACATTTCTGTGTAGACAGAAGTGCNGCTTAAT 480
Db 81 CysValAsnGlyHisAlaGluValValThrPheLeuValAspArgLysCysGlnLeuAsn 100
QY 481 GTCCCTTGATGGCAAGGAGGACACCTCTGATGAAGCTCTACAAATGCGAGAGGAGCT 540
Db 101 ValLeuAspGlyGluGlyArgThrProLeuMetLysAlaLeuGlnCysGluArgGluAla 120
QY 541 TTGTGCAATATTCTCATAGATCGTGGTCTGATCTTAATATTATGTAGATGTGTATGGCA 600
Db 121 -CysAlaAsnIleLeuIleAspAlaGlyAlaAspLeuAsnTyrValAspValTyrGlyAs 140
QY 601 CACGGCTCTCCATTATGCTCCCTTTATAGTGAGAAATTTTAAATGGTGGCAACACTGCTGC 660
Db 140 nThrAlaLeuHisTyrAlaValTyrSerGluAsnLeuLeuMetValAlaThrLeuLeuSe 160
QY 661 CTATGGTGAAGTATCATAGGTGCAAAACAGGCTAGCCCTCACACCCCTTTTACTGGCCAT 720
Db 160 rTyrGlyAlaValIleGluValGlnAsnLysAlaSerLeuThrProLeuLeuAlaAla 180
QY 721 ACAGAAAGAGCAAGCAAACTGTGGAATTTTACTTAACAAAAATGCAAAATGCAAAAGC 780
Db 180 eGlnLysArgSerLysGlnThrValGluPheLeuLeuThrLysAsnAlaAsnAla 200

QY 781 ATTTAATCAGTCTAAATGCACAGCCCTCATGCTTGCCATATGTGAAGGCTCATCAGAGAT 840
Db 200 aPheAsnGluSerLysCysThrAlaLeuMetLeuAlaIleCysGluGlySerGluIle 220
QY 841 AGTCGGATGCTTCTTCTCAGCAAAATGTTGAGCTTGTCTGAAGACATATCAGTAATAAC 900
Db 220 eValGlyMetLeuLeuGlnAsnValAspValPheAlaGluAspIleHisGlyIleThr 240
QY 901 TGCAGAAAGTTATGCTGCTGCTGAGTTAATTACATTCATCAACAACATTTTGAACA 960
Db 240 rAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIleHisGlnLeuLeuGluHis 260
QY 961 TATACGAAAAATTACCTAAATCTCAAAATACCAATCCAGAGAGGAAACATCTCAGGAAC 1020
Db 260 sIleArgLysLeuProLysAsnProGlnAsnThrAsnProGluGlyThrSerThrGlyTh 280
QY 1021 ACCTGATCAGGCTGACCCCTTGGCGGAAAGAACACCTGACACGCTGAAAGCTTGCTGA 1080
Db 280 rProAspGluAlaAlaProLeuAlaGluArgThrProAspThrAlaGluSerLeuLeuG 300
QY 1081 AAAAACACCTGACGAGGCTGCACGCTTGGTGGAGGAAACCTCTGCCAAAATTCATGTCT 1140
Db 300 uLysThrProAspGluAlaAlaArgLeuValGluGlyThrSerAlaLysIleGlnCysLe 320
QY 1141 GGGGAAAGCAACATCTGAAAGTTTGAACAGTCAACAGAGAAACACCTAGGAAATTTT 1200
Db 320 uGlyLysAlaThrSerGlyLysPheGluGlnSerThrGluGluThrProArgLysIleLe 340
QY 1201 GAGGCTTACAAAAGAAACATCTGAGAAATTTTTCATGCCCCAGCAAAAGAAAGATCTAGGA 1260
Db 340 uArgProThrLysGluThrSerGluLysPheSerTrpProAlaLysGluArgSerArgly 360
QY 1261 GATCAGATGGGAGGAAAAAGAAACATCTGTAAGACTGAATGGTGGCAGAGTAACACC 1320
Db 360 sIleThrTrpGluGluLysGluThrSerValLysThrGluCysValAlaGlyValThrPr 380
QY 1321 TAATAAACTCAAGTTTTGAAAAAGAAACATCTAATATGATGTCATGTCTTCAAAAAGA 1380
Db 380 AsnLysThrGluValLeuGluLysGlyThrSerAsnMetIleAlaCysProThrLysG 400
QY 1381 AACATCTACAAAAGCAAGTACAAATGTGATGTGAGTCTGTGAGGCTTATATCTAGTCT 1440
Db 400 uThrSerThrLysAlaSerThrAsnValAspValSerSerValGluProIlePheSerLe 420
QY 1441 TTTTGGCACACGGACTATTGAAATTCACAGTGTACAAAAGTTGAGGAAGACTTTAATCT 1500
Db 420 uPheGlyThrArgThrIleGluAsnSerGlnCysThrLysValGluGluAspPheAsnLe 440
QY 1501 TGCTACCAAGATTATCTTAAGAGTGTGCACAGAATTATACGTGTTTACCTCATGCTAC 1560
Db 440 uAlaThrLysIleIleSerLysSerAlaAlaGlnAsnTyrThrCysLeuProAspAlaTh 460
QY 1561 ATATCAAAAAGATATCAAAACATAATATCAAAATAGAGATCAGATGTCCCATCAGA 1620
Db 460 rTyrGlnLysAspIleLysThrIleAsnHisLysIleGluAspGlnMetPheProSerG 480
QY 1621 ATCCAAAACGAGAGGAGATGAAGAATATTCTTGGATTCTGGGAGTCTCTTTCAGAGTTC 1680
Db 480 uSerLysArgGluGluAspGluGluTyrSerTrpAspSerGlySerLeuPheGluSerSe 500
QY 1681 TGCAAAAGACTCAAGTGTGTATCTGAGTCTATGTATCAGAAAGATAGAGATAAATAG 1740
Db 500 rAlaLysThrGlnValCysIleProGluSerMetTyrGlnLysValMetGluIleAsnAr 520
QY 1741 AGAAGTAGAAGAGGCTTCTCTGAGAGCCATCTGCTTCAAGCTGCGCGTGAATGCAAAA 1800
Db 520 gGluValGluGluLeuProGluLysProSerAlaPheLysProAlaIleGluMetGlnL 540
QY 1801 GACTGTTCCAAATTAAGCCTTTGAATTTGAAGATGAACAAACATTGAGACCACTCAGAT 1860
Db 540 sThrValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAlaGlnMe 560

QY 1861 GTTCCATCAGATCCAAACAAAGGACGATGAAGAAATTTCTGGATTCTGAGAGTCC 1920
 Db 560 tPheProSerGluSerLysGlnLysAspGluGluAsnSerTrpAspSerGluSerPr 580
 QY 1921 CTGTGAGACGGTTTCACAGAAGATGTGATTTACCCAAAGCTACACATCAAAAGAAAT 1980
 Db 580 oCysGluThrValSerGlnLysAspValTrpLysLeuProLysAlaThrHisGlnLysGluPh 600
 QY 1981 CGATACCTTAAGTGGAAATTAAGAGAGTCTCTGTTAAAGATGGTCTTCTGAGCCCTAC 2040
 Db 600 eAspThrLeuSerGlyLysLeuGluGluSerProValLysAspGlyLeuLeuLysProTh 620
 QY 2041 CTGTGCAAGGAAATTTCTTCCAAATAAAGCCTTAGAATTAAGAGGACAGAGAAACATT 2100
 Db 620 rCysGlyArgLysValSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPh 640
 QY 2101 CAAAGCAGAGTCTCTGATAAAGATGGTCTCTGAGCCCTACTCTGGAAGGAAAGTTTC 2160
 Db 640 eLysAlaGluSerProAspLysAspGlyLeuLeuLysProThrCysGlyArgLysValSe 660
 QY 2161 TCTTCCAAATAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGCAGAGTCTCTCTGA 2220
 Db 660 rLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrLeuLysAlaGluSerProAs 680
 QY 2221 TAATGATGGTCTTCTGAAGCCTACTCTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCTTT 2280
 Db 680 pAsnAspGlyLeuLeuLysProThrCysGlyArgLysValSerLeuProAsnLysAlaLe 700
 QY 2281 AGAATTCAGGACAGAGAAACACTCAAGCAGCTCAGATGTTCCCATCAGATCCAAACA 2340
 Db 700 uGluLeuLysAspArgGluThrPheLysAlaAlaGlnMetPheProSerGluSerLysGl 720
 QY 2341 AAAGATGATGAAGAAATTTCTGGGATTTTGAGAGTTTCTCTTGAGACTCTCTTACAGAA 2400
 Db 720 nLysAspAspGluGluAsnSerTrpAspPheGluSerPheLeuGluThrLeuLeuGlnAs 740
 QY 2401 TGATGTGTGTACCAAGGCTACATCAAAAGAAATTCGATACCTTAAGTGGAAATTT 2460
 Db 740 nAspValCysLeuProLysAlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLe 760
 QY 2461 AGAGAGTCTCTGATAAGATGCTCTCTGAGCCTACTCTGGAATGAAATTTCTCT 2520
 Db 760 uGluGluSerProAspLysAspGlyLeuLeuLysProThrCysGlyMetLysIleSerIe 780
 QY 2521 TCCAAATAAAGCCTTAGAATTTGAAGGACAGAGAAACATTCAAGCAGAGAGATGTGAGTTC 2580
 Db 780 uProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaGluAspValSerSe 800
 QY 2581 TGTAGATCCACATTCAGTCTCTTTGGCAACCGACTACTGAAATTCACAGTCTACAA 2640
 Db 800 rValGluSerThrPheSerLeuPheGlyLysProThrThrGluAsnSerGlnSerThrly 820
 QY 2641 AGTTGAGGAGACTTTAAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACCTGGACA 2700
 Db 820 sValGluGluAspPheAsnLeuThrThrLysGluGlyAlaThrLysThrValThrGlyGl 840
 QY 2701 ACAGGAACGTGATATTGGCATTTATTGAACGAGCTCCACAAAGATCAAAACAAATAAGATGCC 2760
 Db 840 nGlnGluArgAspIleGlyIleLeuGluArgAlaProGlnAspGlnThrAsnLysMetPr 860
 QY 2761 CACATCAGATTAAGGAGAAAGAGATACAAATCAAACTTCAGATTCTGAGATTATCTC 2820
 Db 860 oThrSerGluLeuGlyArgLysGluAspThrLysSerThrSerAspSerGluIleIleSe 880
 QY 2821 TGTGAGTGATACAGAAATTTATGAGTGTGTTTACCTGAGGCTACATATCAAAAGAAATAAA 2880
 Db 880 rValSerAspThrGlnAsnTrpGluCysLeuProGluAlaThrTyGlnLysGluIlely 900
 QY 2881 GACAAACAAATGGCAAAATAGAGAGTCTCTGAAAGCCTTCTCACTTTGAGCCTGCCAC 2940
 Db 900 sThrThrAsnGlyLysIleGluSerProGluLysProSerHisPheGluProAlaTh 920
 QY 2941 TGAATGCAAAACTCTGTTCTCCAAATAAAGCCTTAGAATGGAAGAAATAAAACAAACATTGAG 3000

Db 920 rGluMetGlnAsnSerValProAsnLysGlyLeuGluTrpLysAsnLysGlnThrLeuAr 940
 QY 3001 AGCAGATTCAACTACCTTATCAAAATCTTTGGATGCACTTCTCTTGTGAAAGAGAAAG 3060
 Db 940 gAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeuProSerCysGluArgGlyAr 960
 QY 3061 GGAACCTTAAAGAGATTAACCTGACCAAAATACAGCAAAATGCAACAAATGAAATATAA 3120
 Db 960 gGluLeuLysLysAspAsnCysGluGlnIleThrAlaLysMetGluGlnMetLysAsnly 980
 QY 3121 GTTTTGTGTACTACAAAGGAACTGTGCAAGCGAAAGAAATAAAATACAGTTTAGAGAA 3180
 Db 980 sPheCysValLeuGlnLysGluLeuSerGluAlaLysGluIleLysSerGlnLeuGluAs 1000
 QY 3181 CCRAAAAGCTTAAATGGGAACAAGAGCTCTGCGATGTCAGATTGCTTTAAATCAAGAGA 3240
 Db 1000 nGlnLysAlaLysTrpGluGlnLysCysSerValArgLeuThrLeuAsnGlnGluGl 1020
 QY 3241 AGAGAAGAGAAATCTCGATATATTAATAAGAAAAAATTAGACCCCAAGACAACTTAG 3300
 Db 1020 uGluLysArgArgAsnValAspIleLeuLysGluLysIleArgProGluGluGlnLeuAr 1040
 QY 3301 GAAAGATTAGATGAAACACCAACTTGAAACAGACTCTCAGATATCAAGATATAGAAAT 3360
 Db 1040 gLysLysLeuGluValLysGlnGlnLeuGluGlnThrLeuArgIleGlnAspIleGluLe 1060
 QY 3361 GAAAGGTGTAACAAGTAATTTGAATCAGGTTTCTCACACTCATGAAAGTCAAAATGATCT 3420
 Db 1060 uLysSerValThrSerAsnLeuAsnGlnValSerHisThrHisGluSerGluAsnAspLe 1080
 QY 3421 CTTTCATCAAAATTCATGTTGAAAAAGGAAATTCATGCTATAAACTGGAAGTAGCCAC 3480
 Db 1080 uPheHisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluValAlaTh 1100
 QY 3481 ACTGAACATCAACACAGGCTGAGGAAATTAATTAATCTTTGAGCACAATTAAGATTTTACA 3540
 Db 1100 rLeuLysHisGlnHisGlnValLysGluAsnLysTrpPheGluAspIleLysIleLeuGl 1120
 QY 3541 ACAAAGAAATCTCAACTTCAAAATGACCCCTAAATCAAACTGAAACAGAAAAACAGTAAACAAAAG 3600
 Db 1120 nGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGlnLysThrValThrLysAr 1140
 QY 3601 GGCATCTCAGTATAGAGAGAGCTTAAAGTTCTCAGCGCAGAGAACAGATGCTGACTTC 3660
 Db 1140 gAlaSerGlnTrpArgGluGlnLeuLysValLeuThrAlaGluAsnThrMetLeuThrSe 1160
 QY 3661 TAAATTGAAGGAA 3673
 Db 1160 rLysLeuLysGlu 1164

RESULT 2

US-10-408-765A-1557
 ; Sequence 1557, Application US/10408765A
 ; Publication No. US20040101874A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE REFERENCE: 660088.465
 ; CURRENT APPLICATION NUMBER: US/10/408,765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1557
 ; LENGTH: 1011
 ; TYPE: PRF

; ORGANISM: Homo sapiens

US-10-408-765A-1557

Alignment Scores:

Pred. No.: 3,16e-302 Length: 1011
Score: 5173.00 Matches: 1011
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.23% Indels: 0
DB: 16 Gaps: 0

US-09-602-362E-26 (1-3673) x US-10-408-765A-1557 (1-1011)

641 ATGGTGGCAACACTGTCTCTATGTGTCAGTCTCATCGAGGTGCAAAACAGGCTAGCCCTC 700
Db 1 MetValAlaThrLeuLeuSerTyrGlyAlaValIleGluValGlnAsnLysAlaSerLeu 20
QY 701 ACACCCCTTTTACTGGCCATACAGAAAGAGCAAGCAAACTGTGGAATTTTACTAACA 760
Db 21 ThrProLeuLeuAlaIleGlnLysArgSerLysGlnThrValGluPheLeuLeuThr 40
QY 761 AAAAATGCAAAATGCAACGCTTTAATAGTCTAAATGACACGCCCTCATGCTTGGCATA 820
Db 41 LysAsnAlaAsnAlaAsnAlaPheAsnGluSerLysCysThrAlaLeuMetLeuAlaIle 60
QY 821 TGTGAAGGCTCATCAGAGATAGTCGGCATGCTCTTCTTCAGCAAAATGTTGACGTCTTTGCT 880
Db 61 CysGluGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAlaPheAla 80
QY 881 GAAGACATACATGAATACTGCAGAACGTTATGCTGCTGCTCGTGGAGTTAATTACATT 940
Db 81 GluAspIleHisGlyIleThrAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIle 100
QY 941 CATCAACAACCTTTTGGAAACATATACGAAATATTACCTTAAATCTCTCAAAATACCAATCCA 1000
Db 101 HisGlnGlnLeuLeuGluHisIleArgLysLeuProLysAsnProGlnAsnThrAsnPro 120
QY 1001 GAAGGACATCTACAGAACACCTGATGAGCTGCACCCCTGCGGAAAGAACACCTGAC 1060
Db 121 GluGlyThrSerThrGlyThrProAspGluAlaAlaProLeuAlaGluArgThrProAsp 140
QY 1061 ACGGCTGAAGAGCTTGTGGAAAAACACCTGACGAGGCTGCAGCTTGTGGAGGGAACG 1120
Db 141 ThrAlaGluSerLeuLeuGluLysThrProAspGluAlaAlaArgLeuValGluGlyThr 160
QY 1121 TCTGCCAAATTCATGCTGGGGAAGCACTGCTGGAAGTTTGAACAGTCAACAGAA 1180
Db 161 SerAlaLysIleGlnCysLeuGlyLysAlaThrSerGlyLysPheGlnSerThrGlu 180
QY 1181 GAAACACCTAGGAAATTTTGGAGCTTACAAAAGAAACATCTGAGAAATTTTCATGGCCA 1240
Db 181 GluThrProArgLysIleLeuArgProThrLysGluThrSerGluLysPheSerTrpPro 200
QY 1241 GCAAAAGAAAGATCTAGGAAGATCATATGGAGGAAAAAGAACATCTCTTAAAGACTGAA 1300
Db 201 AlaLysGluArgSerArgLysIleThrTrpGluGluLysGluThrSerValLysThrGlu 220
QY 1301 TGGGTGCGAGGATTAACACTATAAACTGAAGTTTGGAAAAAGAACATCTAATATG 1360
Db 221 CysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThrSerAsnMet 240
QY 1361 ATTGCATGCTCTCAAAAGAAACATCTACAAAAGCAAGTACAAATCGATGTGAGTTCT 1420
Db 241 IleAlaCysProThrLysGluThrSerThrLysAlaSerThrAsnValAlaPheValSerSer 260
QY 1421 GTAGAGCTTATATCAGTCTTTTGGCACCGACTATTGAAATTTCAAGTGTACAAA 1480
Db 261 ValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGlnCysThrLys 280
QY 1481 GTTGAGGAAGACTTTAATCTTGCTACCAAGATTATCTCTAAGAGTCTCTCCACAGAAATTAT 1540
Db 281 ValGluGluAspPheAsnLeuAlaThrLysIleIleSerLysSerAlaAlaGlnAsnTyr 300

QY 1541 ACGTGTTTTACCTGATGCTACATATCAAAAAGATATCAAAAACAATATAATCAAAAATAGAA 1600
Db 301 ThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHisLysIleGlu 320
QY 1601 GATCAGATGTTCCCATCAGATCCAAACGAGAGAGATCAAGANATATCTTGGGATTC 1660
Db 321 AspGlnMetPheProSerGluSerLysArgGluGluAspGluGluThrSerTrpAspSer 340
QY 1661 GCGAGTCTCTTTGAGAGTCTTGCAAAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAG 1720
Db 341 GlySerLeuPheGluSerSerAlaLysThrGlnValCysIlePheGluSerMetTyrGln 360
QY 1721 AAAGTAATGGAGATAAATAGAGAGATAGAGAGCTTCTCTGAGAAGCCATCTGCTTCAAG 1780
Db 361 LysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSerAlaPheLys 380
QY 1781 CCTCCGTNGAAATGCAAAAGACTGTTCCAAATAAAGCCTTTCATTTGAAGAAATGAACAA 1840
Db 381 ProAlaValGluMetGlnLysThrValProAsnLysAlaPheGluLeuLysAsnGluGln 400
QY 1841 ACATTGAGAGCAGCTCAGATGTTCCCATCAGATCCAAAACAAAGGACGATGAAGAAAT 1900
Db 401 ThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAspGluGluAsn 420
QY 1901 TCTTGGCATCTGAGAGTCCCTGTGACACGTTTCACAGAGGATGTGTATTACCCAAA 1960
Db 421 SerTrpAspSerGluSerProCysGluThrValSerGlnLysAspValTyrLeuProLys 440
QY 1961 GCTACACATCAAAAAGAAATTCGATACCTTAAAGTGGAAAAATTAGAGAGTCTCTGTGTA 2020
Db 441 AlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGluSerProValLys 460
QY 2021 CATGCTCTCTGAGACCTACTCTGGAAGAAAGATTTCTCTCCAAATAAAGCCTTAGAA 2080
Db 461 AspGlyLeuLeuLysProThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGlu 480
QY 2081 TTAAGGACACAGAGAAACATTCAAAGCAGAGTCTCTGATAAAGATGCTCTTGAAGCCT 2140
Db 481 LeuLysAspArgGluThrPheLysAlaGluSerProAspLysAspGlyLeuLeuLysPro 500
QY 2141 ACCTGTGAGAGAAAGTTCTCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACA 2200
Db 501 ThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThr 520
QY 2201 CTCAAAGCAGAGTCTCTGATATATGATGGTCTCTGAAAGCCTACCTGTGGAAGAAAGTT 2260
Db 521 LeuLysAlaGluSerProAspAsnAspGlyLeuLeuLysProThrCysGlyArgLysVal 540
QY 2261 TCTCTTCCAAATAAAGCTTTAGAAATTAAGGACAGAGAAACATTCAAAGCAGCTCAGATG 2320
Db 541 SerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaAlaGlnMet 560
QY 2321 TTCCCATCAGATCCAAACAAAGGATGATGAGAAATTTCTGGGATTTTGGAGTTTC 2380
Db 561 PheProSerGluSerLysGlnLysAspGluGluAsnSerTrpAspPheGluSerPhe 580
QY 2381 CTTGAGACTCTCTTACAGAATGATGTGTTTACCACCAAGGCTACACATCAAAAGAAATTC 2440
Db 581 LeuGluThrLeuLeuGlnAsnAspValCysLeuProLysAlaThrHisGlnLysGluPhe 600
QY 2441 GATACCTTAAGTGAAATTTAGAGAGTCTCTGATAAAGATGGTCTCTGAGCCTACC 2500
Db 601 AspThrLeuSerGlyLysLeuGluGluSerProAspLysAspGlyLeuLeuLysProThr 620
QY 2501 TGTGGAATGAAATTTCTCTTCCAAATAAAGCCTTTAGAAATTAAGGACAGAGAAACATTC 2560
Db 621 CysGlyMetLysIleSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPhe 640
QY 2561 AAAGCAGAGATGTGAGTCTCTGATAGTCCACATTCAGTCTTTTGGCAACCGACTACT 2620
Db 641 LysAlaGluAspValSerSerValGluSerThrPheSerLeuPheGlyLysProThrThr 660
QY 2621 GAAATTCACAGTCTCAAAAAGTTGAGGAGAGCTTTAATCTTACTCAACAGGAGGAGCA 2680

Db	661	GlusAsnSerGlnSerThrLysValGluGluAspPheAsnLeuThrThrLysGluGlyAla	680
QY	2681	ACAAAGACAGTAACCTGGCAACAGACAGCTGTATATGGCATTTATGACGAGCTCCACAA	2740
Db	681	ThrLysThrValThrGlyGlnGlnGluArgAspIleGlyIleIleGluArgAlaProGln	700
QY	2741	GATCAAAACAAATAGATGCCACATCAGAAATAGGAAGAAAGAGATACAAAATCAACT	2800
Db	701	AspGlnThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThr	720
QY	2801	TCAGATTCTGAGATTATCTCTGTGAGTGATACACAGAAATTATGAGTGTATTACCTGAGCT	2860
Db	721	SerAspSerGluIleSerValSerAspThrGlnAsnTyrgluCysLeuProGluAla	740
QY	2861	ACATATCAAAAAGAAATAAGACAAACAAATGCGAAATAGAGAGCTCTCCTGAAAGGCT	2920
Db	741	ThrTyrglnLysGluIleLysThrThrAsnGlyLysIleGluGluSerProGluLysPro	760
QY	2921	TCTCACTTTGAGCTGCCACTGAAATGCAAACTCTGTTTCCAAATAAAGGCTTAGAATGG	2980
Db	761	SerHisPheGluProAlaThrGluMetGlnAsnSerValProAsnLysGlyLeuGluTrp	780
QY	2981	AGAATAAACAAACATTGAGACAGATTCAACTACCTATCAAAATCTTGGATGCACTT	3040
Db	781	LysAsnLysGlnThrLeuArgAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeu	800
QY	3041	CCTCTTGTGAAGAGGAAGGAACTTAAAGAGATAACTGTGAACAAATACAGCAAAA	3100
Db	801	ProSerCysGluArgGlyArgGluLeuLysAspAsnCysGluGlnIleThrAlalys	820
QY	3101	ATGGAACAAATGAAATAAAGTTTGTGTACTACAAAGAACTGTACAGACGGAAGAA	3160
Db	821	MetGluGlnMetLysAsnLysPheCysValLeuGlnLysGluLeuSerGluAlalysGlu	840
QY	3161	ATAAAATCACAGTTAGAGAACCAAAAGCTAAATGGGAACAGAGCTCTGAGTGTGAGA	3220
Db	841	IleLysSerGlnLeuGluAsnGlnLysAlalysTrpGluGlnLeuLysSerValarg	860
QY	3221	TTGCCCTTAAATCAAGAAGAGAGAGAGAAATGTCGATATATTAAGAAAAAATTT	3280
Db	861	LeuProLeuAsnGlnGluGluLysArgAsnValAspIleLeuLysGluLysIle	880
QY	3281	AGACCCGAGAGCAACTTAGGAAAAGTTAGAAGTGAACACCACTTGAACAGACTCTC	3340
Db	881	ArgProGluGlnLeuArgLysLysLeuGluValLysHisGlnLeuGluGlnThrLeu	900
QY	3341	AGAATAACAAGATATAGAATTGAAAAGTGPAAACAAGTAATTTCAATCAGGTTTCTC	3400
Db	901	ArgIleGlnAspIleGluLysSerValThrSerAsnLeuAsnGlnValSerHisThr	920
QY	3401	CATGAAGTGAAATGATCTCTTCATGAATTCGATGTTGAAAAGGAAATTCGCCATG	3460
Db	921	HisGluSerGluAsnAspLeuPheHisGluAsnCysMetLeuLysGluIleAlaMet	940
QY	3461	CTAAAACTGGAAGTAGCCACACTGAAACATCAACACAGCTCAAGGAAATAAATAC	3520
Db	941	LeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsnLysTyrrPhe	960
QY	3521	GAGCAATTAAGATTTCAGAAAGAATGCTGAACCTCAATGACCCATAAATGAAA	3580
Db	961	GluAspIleLysIleLeuGlnLysAsnAlaGluLeuGlnMetThrLeuLysLeuLys	980
QY	3581	CAGAAAACAGTAACAAAAAGGGCATCTCAGTATAGAGACAGCTTAAAGTCTCGACG	3640
Db	981	GlnLysThrValThrLysArgAlaSerGlnTyrrArgGluGlnLeuLysValLeuThrAla	1000
QY	3641	GAGAACACAGATGCTACTTCTAAATTAAGGAA 3673	
Db	1001	GlusAsnThrMetLeuThrSerLysLeuLysGlu 1011	

RESULT 3
US-10-007-805-577

; Sequence 577, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yudiu
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 577
; LENGTH: 1239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-577

Alignment Scores:
Pred. No.: 3,24e-302 Length: 1239
Score: 5173.00 Matches: 1011
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.23% Indels: 0
DB: 13 Gaps: 0

US-09-602-362E-26 (1-3673) x US-10-007-805-577 (1-1239)

QY	641	ATGTTGGCAACACACGTCTCTATGTTGTCAGTCACTCAGCTGCAAAAACAGGCTAGCCTC	700
Db	1	MetValAlaThrLeuLeuSerTyrrGlyAlaValIleGluValGlnAsnLysAlaSerLeu	20
QY	701	ACACCCCTTTTACTGGCCATACAGAAAAGAACGAAACCTGTGGAATTTTACTAACA	760
Db	21	ThrProLeuLeuAlaIleGlnLysArgSerLysGlnThrValGluPheLeuLeuThr	40
QY	761	AAAAATCGAATGCAACGCACTTTAATGAGTCTAAATGACAGCCCTCATGCTTGCCTA	820
Db	41	LysAsnAlaAsnAlaAsnAlaPheAsnGluSerLysCysThrAlaLeuMetLeuAlaIle	60
QY	821	TGTCAAGGCTCATCAGAGATAGTCGGCATGCTTCTTCCAGCAAAATGTTGACGCTTTT	880
Db	61	CysGluGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAla	80
QY	881	GAAGACATACATGGAATTAACGTGAGAACGTTATGCTGCTGCTGAGTGAATTAATTA	940
Db	81	GluAspIleHisGlyIleThrAlaGluArgTyrrAlaAlaAlaArgGlyValAsnTyrrIle	100
QY	941	CATCAACAACTTTTGGACATATACGAAATTAACATAAAATCTCAAAATACCAATCCA	1000
Db	101	HisGlnGlnLeuLeuGluHisIleArgLysLeuProLysAsnProGlnAsnThrAsnPro	120
QY	1001	GAAGAACATCTACAGAAACACCTGATGAGCTGCACCTTGGCGGAAAGAACACCTGAC	1060
Db	121	GluGlyThrSerThrGlyThrProAspGluAlaAlaProLeuAlaGluArgThrProAsp	140
QY	1061	ACGGCTGAAGCTGTGGAAAAACACCTGACGAGGCTCAGCGCTTGGTGGGGGAAACG	1120
Db	141	ThrAlaGluSerLeuLeuGluLysThrProAspGluAlaAlaArgLeuValGluGlyThr	160
QY	1121	TCTGCCAAAATTCATCTCTGGGGAAGAACACATCTCGAAAGTTTGAACTGACACAGAA	1180
Db	161	SerAlaLysIleGlnCysLeuGlyLysAlaThrSerGlyLysPheGluGlnSerThrGlu	180

1181 GAAACACCTAGGAAATTTTGGGCTCACAAGAAACATCTCAGAAATTTTCATGGCCA 1240
Db GluThrProArgLysIleLeuArgProThrLysGluThrSerGluLysPheSerTrpPro 200
1241 GCAAAAGAAAGATCTAGGAGATCATGCGAGGAGAAAGAAACATCTCTAAGACTGAA 1300
Db AlaLysGluArgSerArgLysIleThrTrpGluLysGluThrSerValLysThrGlu 220
1301 TGGGTGCAGAGTAACACCTAATAAAGTCTTGAAGTTTGGAAAGAAACATCTAATATG 1360
Qy CysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThrSerAsnMet 240
1361 ATTGCATGTCCTCAAAAGAAACATCTACAAGACAGTACAAATGTGATGTGATGTTCT 1420
Db IleAlaCysProThrLysGluThrSerThrLysAlaSerThrAsnValAspValSerSer 260
1421 GTAGAGCCTATATTCAGTCTTTTGGCACACGGAATATTGAAATTCACAGTGTACAAA 1480
Qy ValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGlnCysThrLys 280
1481 GTTGAGGAAGACTTAACTTGTGTACCAAGATTATCTTAAGAGTCTGCACAGAAATTAT 1540
Qy ValGluGluAspPheAsnLeuAlaThrLysIleSerLysSerAlaAlaGlnAsnTyr 300
1541 ACCTGTTTACCTGATCATATCAATCAAAAGATATCAAAACATAAATCACAAAATAGAA 1600
Db ThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHisLysIleGlu 320
1601 GATCAGATGTTCCATCAGAAATCCAAACGAGAGGAAGATGAAGATATCTTGGATCTC 1660
Qy AspGlnMetPheProSerGluSerLysArgGluGluAspGluLysSerTrpAspSer 340
1661 GGAGTCTCTTTCAGAGTCTGCAAGAGCTCAAGTGTGTATCCTCAGTCTATGATATCAG 1720
Db GlySerLeuPheGluSerSerAlaLysThrGlnValCysIleProGluSerMetTyrGln 360
1721 AAAGTAAATGAGATAAATAGAAAGTAGAAGCTTCTCAGAAAGCCATCTCGCTTCAAG 1780
Qy LysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSerAlaPheLys 380
1781 CTTGCCGTGAAATGCAAGAGCTGTCCAAATTAAGCCCTTCAATTGAAGAAATGAACAA 1840
Db ProAlaValGluMetGlnLysThrValProAsnLysAlaPheGluLeuLysAsnGluGln 400
1841 ACATTGAGAGCAGCTCAGATGTTTCCCATCAGAAATCCAAACAAAGACCATGAAGAAAT 1900
Qy ThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAspGluGluAsn 420
1901 TCTTGGATTCAGAGTCCCTGTGAGAGGTTTACAGAAGATGTATTATCCCAA 1960
Db SerTrpAspSerGluSerProCysGluThrValSerGlnLysAspValTyrLeuProLys 440
1961 GCTACACATCAAAAGAAATTCGATACCTTAAGTGGAAATTAGAAGAGTCTCTGTATAA 2020
Qy AlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGluGluSerProValLys 460
2021 GATGTCCTTGAAGCCTACCTGTGAAGAAAGTTTCTCTCCAAATTAAGCCCTTAGAA 2080
Qy AspGlyLeuLeuLysProThrCysGlyValSerLysValSerLeuProAsnLysAlaLeuGlu 480
2081 TTAAGCAGACAGAAACATTCAAAGCAGAGTCTCCTGATAAGATGCTCTCTGAAGCT 2140
Db LeuLysAspArgGluThrPheLysAlaGluSerProAspLysAspGlyLeuLeuLysPro 500
2141 ACCTGCGAAGAAAGTTTCTCTCCAAATAAAGCCTTAGAATTAAAGACAGAGAAACA 2200
Qy ThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGluLysAspArgGluThr 520
2201 CTCAAAGCAGACTCTCTGATATGATGCTTCTGAAGCCCTACCTGTGGAAGGAAGTT 2260
Qy LeuLysAlaGluSerProAspAsnAspGlyLeuLeuLysProThrCysGlyArgLysVal 540
2261 TCTCTTCCAAATAAAGCTTTAGAAATTAAGAGGACAGAGAAACATTCAAAGCAGCTCAGATG 2320

541 SerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaAlaGlnMet 560
2321 TTCCCATCAGAAATCAAAACAAAGGATGATCAAGAAATTTCTGGGATTTTGAGAGTTTC 2380
Qy PheProSerGluSerLysGlnLysAspAspGluGluAsnSerTrpAspPheGluSerPhe 580
2381 CTTGAGACTCTCTTACAGAAATGATGCTCTCTTACCAGAGCTACACATCAAAAGAAATTC 2440
Qy LeuGluThrLeuLeuGlnAsnAspValCysLeuProLysAlaThrHisGlnLysGluPhe 600
2441 GATACCTTAAGTGGAAATTAGAAGAGTCTCTGATAAAGATGCTCTCTGAAAGCCCTACC 2500
Qy AspThrLeuSerGlyLysLeuGluGluSerProAspLysAspGlyLeuLeuLysProThr 620
2501 TGTGGAATGAAATTTCTCTTCCAAATAAAGCCTTAGAATTTGAAGACAGAGAAACATTC 2560
Qy CysGlyMetLysIleSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPhe 640
2561 AAAGCAGAGGATGAGTCTCTGAGCTCCACATTCAGTCTCTTTTGGCAAAACCGACTACT 2620
Qy LysAlaGluAspValSerSerValGluSerThrPheSerLeuPheGlyLysProThrThr 660
2621 GAAATTCACAGTCTACAAAGTTGAGGAAGACTTAAATCTTACTACCAAGGAGGAGCA 2680
Qy GluAsnSerGlnSerThrLysValGluGluAspPheAsnLeuThrThrLysGluGlyAla 680
2681 ACAAGACAGTACTGACACCAACGAGACGTGATATTGGCATTATTGACAGCTCCACAA 2740
Qy ThrLysThrValThrGlyGlnGlnArgAspIleGlyIleLeuGluArgAlaProGln 700
2741 GATCAAAACAAATAGATGCCCATCAGATCAGAAATTAGGAAGAAAGAAATACAAATCAACT 2800
Qy AspGlnThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThr 720
2801 TCAGATTCTGAGATTATCTCTGTGAGTGATACACAGAAATATGATGTTTACCTGAGGCT 2860
Qy SerAspSerGluIleIleSerValSerAspThrGlnAsnTyrGluCysLeuProGluAla 740
2861 ACATATCAAAAGAAATTAAGACAAACAAATAGCAAAATAGAAAGTCTCTCTGAAAGCCCT 2920
Qy ThrTyrGlnLysGluIleLysThrThrAsnGlyLysIleGluGluSerProGluLysPro 760
2921 TCTCATTTGAGCTGCTGCACTGAAATCCAAACTCTGTTCCAAATTAAGCCTTAGAATGG 2980
Qy SerHisPheGluProAlaThrGluMetGlnAsnSerValProAsnLysGlyLeuGluTrp 780
2981 AAGAATAACAAACATTTGAGAGCAGATTCAACTACCTATCAAAATCTTTGGATGCACTT 3040
Qy LysAsnLysGlnThrLeuArgAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeu 800
3041 CTTTCTTTGTCAAGAGGAGGAACTTTAAATAAGATAACTGTGAACAAATTAAGCAAAA 3100
Qy ProSerCysGluArgGlyArgGluLeuLysLysAspAsnCysGluGlnIleThrAlaLys 820
3101 ATGGAACAAATGAAATTAAGTTTGTGTACTACAAAGGAACCTGTGAGAGGGAAGAA 3160
Qy MetGluGlnMetLysAsnLysPheCysValLeuGlnLysGluLeuSerGluAlaLysGlu 840
3161 ATAAATCAGATTAGAGAACCAAAAGCTTAAATGGGAACAGAGCTCTGCACTGTGAGA 3220
Qy IleLysSerGlnLeuGluAsnGlnLysAlaLysTrpGluGlnGluLeuCysSerValArg 860
3221 TTGCTTTTAATCAGAGAGAGAGAGAGAAATGTGATATATAAGAGAAATTAAGAGAAAT 3280
Qy LeuProLeuAsnGlnGluGluLysArgAsnValAspIleLeuLysGlyLysIle 880
3281 AGACCCCAAGAGCAACTTAGCAAAATTTAGAAAGTGAACCAACCAACTTGAACAGACTCTC 3340
Qy ArgProGluGluGlnLeuArgLysLysLeuGluValLysHisGlnLeuGluGlnThrLeu 900
3341 AGAATCAAGATATAGAAATGAAAGTGAACAGTAATTTGAATCAGGTTTCTCACACT 3400

Db 901 ArgileGlnAspIleuLeuLysSerValThrSerAsnLeuAsnGlnValSerHisThr 920
QY 3401 CATGAAGTGAATGATCTCTTTTCATGAAATTCGATGTTGAAAGGAAATGCCCATG 3460
Db 921 HisGluSerGluAsnAspLeuPheHisGluAsnCysMetLeuLysGluIleAlanMet 940
QY 3461 CTAAACTGGAAGTAGCCACACTGAAACATCATCAACACGAGGTGAAGGAAATTAATACTTT 3520
Db 941 LeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsnLysTyrPhe 960
QY 3521 GAGGACATTAAAGATTTCAGAAAGAAAGATGCTGAACCTTCAATGACCTTAAACTGAAA 3580
Db 961 GluAspIleLysIleLeuGlnLysAsnAlaGluLeuMetThrLeuLysLeuLys 980
QY 3581 CAGAAAACAGTAAACAAAGGGCATCTCAGTATAGAGACGACTTAAAGTCTGACGGCA 3640
Db 981 GlnLysThrValThrLysArgAlaSerGlnTyrArgGluGlnLeuLysValLeuThrAla 1000
QY 3641 GAGAACCATGCTGACTTCTTAAATGAAGAA 3673
Db 1001 GluAsnThrMetLeuThrSerLysLeuLysGlu 1011
RESULT 4
US-10-076-622-577
; Sequence 577, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 577
; LENGTH: 1239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-076-622-577

Alignment Scores:
Pred. No.: 3-24e-302 Length: 1239
Score: 5173.00 Matches: 1011
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.23% Indels: 0
DB: 14 Gaps: 0

US-09-602-362E-26 (1-3673) x US-10-076-622-577 (1-1239)

QY 641 ATGGTGGCAACACTGCTGCTCTATGTCGAGTCATCGAGGTGCAAAACAGAGCTAGCCTC 700
Db 1 MetValAlaThrLeuLeuSerTyrGlyAlaValIleGluValGlnAsnLysAlaSerLeu 20
QY 701 ACACCCCTTTTACTGGCCCATACAGAAAGAAAGCAAGCAAACTGTGGAATTTTACTAACA 760
Db 21 ThrProLeuLeuLeuAlaIleGlnLysArgSerLysGlnThrValGluPheLeuLeuThr 40
QY 761 AAAAATGCAAAATGCAACCGATTAAATGATCTTAAATGACACGCCCTCATGCTGCGATA 820
Db 41 LysAsnAlaAsnAlaPheAsnGluSerLysCysThrAlaLeuMetLeuAlaIle 60
QY 821 TGTGAAGGCTCATCAGAGTAGTCGCGATGCTTCTTTCAGCAAAATGTTGACGCTTTTGTCT 880
Db 61 CysGluGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAla 80
QY 881 GAAGACATACATGGAATAACTGAGAACGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 940
Db 81 GluAspIleHisGlyIleThrAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIle 100

QY 941 CATCAACACATTTTGGAAACATATACGAAATATTACCTTAAATCTCTCAAAATACCAATCCA 1000
Db HisGlnGlnLeuLeuGluHisIleArgLysLeuProLysAsnProGlnAsnThrAsnPro 120
QY 1001 GAAGGAACATCTCAGGAACACACTGATGAGCTGCACCTTGGCGGAAAGAACACCTGAC 1060
Db 121 GluGlyThrSerThrGlyThrProAspGluAlaAlaProLeuAlaGluArgThrProAsp 140
QY 1061 ACGCTGAAAGCTTGTGGAAAAACACCTGACAGAGCTGCAGCTTGTGGTGGAGGAAACG 1120
Db 141 ThrAlaGluSerLeuLeuGluLysThrProAspGluAlaAlaArgLeuValGluGlyThr 160
QY 1121 TCTGCCAAAATTCATCTCTGGGGAAGCAACATCTGGAAGTTTGAACAGTCAACAGAA 1180
Db 161 SerAlaLysIleGlnCysLeuGlyLysAlaThrSerGlyLysPheGluGlnSerThrGlu 180
QY 1181 GAAACACCTAGGAAAATTTTGAAGCCTPACAAAAGAAACATCTGAGAAATTTTCATGCCCA 1240
Db 181 GluThrProArgLysIleLeuArgProThrLysGluThrSerGluLysPheSerTyrPro 200
QY 1241 GCAAAAGAAAGATCTAGGAGATCAGATGGAGGAAAGAAACATCTGTAAAGACTGAA 1300
Db 201 AlaLysGluArgSerArgLysIleThrTyrGluGluLysGluThrSerValLysThrGlu 220
QY 1301 TGCCTGGCAGAGTAACACACCTTAATAAACTCAAGTTTGTGAAAAGGAAACATCTAATATG 1360
Db 221 CysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThrSerAsnMet 240
QY 1361 ATTGCATGCTCTCAAAAGAAACATCTACAAAGCAAGTACAAATGTGGATGTGAGTTCT 1420
Db 241 IleAlaCysProThrLysGluThrSerThrLysAlaSerThrAsnValAspValSerSer 260
QY 1421 GTAGAGCTTATTCAGTCTTTTGGCACACGAGTATTCAAATTCACAGTGTACAAAA 1480
Db 261 ValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGlnCysThrLys 280
QY 1481 GTTGAGGAAGACTTTAATCTTGTCTACCAAGATTATCTTAAGAGTGTCTGACAGAAATAT 1540
Db 281 ValGluGluAspPheAsnLeuAlaThrLysIleSerLysSerAlaAlaGlnAsnTyr 300
QY 1541 ACGTGTTTTACCTGATGCTACATATCAAAAGATATCAAAACATTAATCAAAATAGAA 1600
Db 301 ThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHisLysIleGlu 320
QY 1601 GATCAGATGTTCCCATCAGAAATCCAAACGAGAGAGAGTTCCTGAGAGCCATCTGCTTCAAG 1660
Db 321 AspGlnMetPheProSerGluSerLysArgGluGluAspGluGluTyrSerTyrAspSer 340
QY 1661 GGGAGTCTCTTTGAGAGTCTGCAAGACTCAAGTGTGTATACCTGAGTCTATCTATCAG 1720
Db 341 GlySerLeuPheGluSerSerAlaLysThrGlnValCysIleProGluSerMetTyrGln 360
QY 1721 AAAGTAATGGAGATAAATAGAGAGTAGAAGTTCCTGAGAGCCATCTGCTTCAAG 1780
Db 361 LysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSerAlaPheLys 380
QY 1781 CCTCCCTGNGAAATGCAAAAGACTGTTCCAAANTAAAGCCCTTTGAAATGAAGATGAACAA 1840
Db 381 ProAlaValGluMetGlnLysThrValProAsnLysAlaPheGluLeuLysAsnGluGln 400
QY 1841 ACATTGAGAGCAGCTCAGATGTTCCCATCAGATCCAAACAAAGAGCAGATGAAGAAAT 1900
Db 401 ThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspGluGluAsn 420
QY 1901 TCTTGGGATTCGAGAGTCCCTGTGAGACGCTTTCACAGAGGATGTGTATTTTACCCAAA 1960
Db 421 SerTyrAspSerGluSerProCysGluThrValSerGlnLysAspValTyrLeuProLys 440
QY 1961 GCTACATCAAAAGAAATTCGATACCTTAAGTGGAAATTAGAGATCTCTCTGTTAAA 2020
Db 441 AlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGluGluSerProValLys 460
QY 2021 GATGCTCTTCTGAAGCCTACCTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAA 2080

701 ACACCCCTTTTACTGGCCATACAGAAAAGAACAGCAAACTGTGGAATTTTACTAACA 760
Db
21 ThrProLeuLeuLeuAlaIleGlnLysArgSerLysGlnThrValGluPheLeuLeuThr 40
761 AAAAAATGCAATGCAAAACGCTTTAATAGTCTAAATGCACAGCCCTCATGCTTGCCATA 820
Db
41 LysAsnAlaAsnAlaAsnAlaPheAsnGluSerLysCysThrAlaLeuMetLeuAlaIle 60
821 TGTGAAGGCTCATCAGAGATAGTCGCGCATGCTTCTTCAGCAAAATGTTGAGCTCTTGCT 880
Db
61 CysGluGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAla 80
881 GAAGACATACATGGAATRACTGCAGAACGTTATGCTGCTGCTGCGGAGTTAATACATT 940
Db
81 GluAspIleHisGlyIleThrAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIle 100
941 CATCAACAATTTTGGACATATACGAAAATTTACCTAAATCTCAAAATACCAATCCA 1000
Db
101 HisGlnGlnLeuLeuGluHisIleArgLysLeuP-LysAsnProGlnAsnThrAsnPro 120
1001 GAAGGAACATCTACAGGAACACCTGATGAGGCTGCACCTTGGCGGAAGAACACCTGAC 1060
Db
121 GluGlyThrSerThrGlyThrProAspGluAlaAlaProLeuAlaGluArgThrProAsp 140
1061 ACGGCTGAAGCTTGTGGAAGAAACACCTGACGAGGCTGCACGCTTGGTGGAGGAACG 1120
Db
141 ThrAlaGluSerLeuLeuGluLysThrProAspGluAlaAlaArgLeuValGluGlyThr 160
1121 TCTGCAAAATTCATGCTGGGAAAGCAACATCTGGAAGCTTTCAACAGTCAACAGAA 1180
Db
161 SerAlaLysIleGlnCysLeuGlyLysAlaThrSerGlyLysPheGluGlnSerThrGlu 180
1181 GAAACACCTPAGGAATTTTGGGCTACAAAAGAAACATCTGAGAAATTTTCATGGCCA 1240
Db
181 GluThrProArgLysIleLeuArgProThrLysGluThrSerGluLysPheSerTrpPro 200
1241 GCAAAAGAAATCTAGGAAGATCACATGGGAGGAAAGAAACATCTGTAAGACTGAA 1300
Db
201 AlaLysGluArgSerArgLysIleThrTrpGluGluLysGluThrSerValLysThrGlu 220
1301 TGGTGGCAGGAGTACACCTAATAAACTGAAGTTTGGAAAAAGGAACATCTAATATG 1360
Db
221 CysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThrSerAsnMet 240
1361 ATGCGATGCTTACAAAAGAAACATCTACAAAAGCAAGTACAAATGTGGATGTGATGTTCT 1420
Db
241 IleAlaCysProThrLysGluThrSerThrLysAlaSerThrAsnValAspValSerSer 260
1421 GTAGACCTATATTCAGTCTTTTGGCACACGACTATTGAAAATTCACAGTGTACAAA 1480
Db
261 ValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGlnCysThrLys 280
1481 GTTGAGGAAGACTTTAATCTTGCTACCAAGATTATCTTAAAGATGCTGCACAGAAATTAT 1540
Db
281 ValGluGluAspPheAsnLeuAlaThrLysIleIleSerLysSerAlaAlaGlnAsnTyr 300
1541 ACCTGTTTACCTGATCTACATACAAAAGATATCAAAACATTAATCACAATAATAGAA 1600
Db
301 ThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHisLysIleGlu 320
1601 GATCAGATGTTCCCATCAGAAATCCAAACGAGAGGAGATGAAGAATTAATCTTGGGATCT 1660
Db
321 AspGlnMetPheProSerGluSerLysArgGluGluAspGluGluTyrSerTrpAspSer 340
1661 GGGAGTCTCTTTGAGAGTCTGAAAAGACTCAAGTGTGTATACCTGAGTCTATGATCAG 1720
Db
341 GlySerLeuPheGluSerSerAlaLysThrGlnValCysIleProGluSerMetTyrGln 360
1721 AAAGTAAATGAGATAAATAGAAAGTAGAAGACTTCTCAGAAAGCCATCTGCTCTCAAG 1780
Db
361 LysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSerAlaPheLys 380
1781 CCTGCCGTGAAATGCAAAAGACTGTTCCAAATAAAAGCCTTTGAATTTGAAGAAATGAACA 1840

381 ProAlaValGluMetGlnLysThrValProAsnLysAlaPheGluLeuLysAsnGluGln 400
Qy
1841 ACATTGACAGACCTCAGATGTTCCCATCAGATCCAAACAAAGACGATGAGAAAT 1900
Qy
401 ThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspGluGluAsn 420
Qy
1901 TCTTGGATTCTGAGAGTCCCTGTGAGACGTTTTCAGAAAGGATGTGTATTTACCCAAA 1960
Db
421 SerTrpAspSerGluSerProCysGluThrValSerGlnLysAspValTyrLeuProLys 440
Qy
1961 GCTACACATCAAAAGAAATTCGATACCTTAAGTGGAAAAATTAGAGAGTCTCTGTATAA 2020
Db
441 AlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGluLeuSerProValLys 460
Qy
2021 GATGCTCTTCTGAAGCCTACTCTGTGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAA 2080
Db
461 AspGlyLeuLeuLysProThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGlu 480
Qy
2081 TTAAGGACAGAGAAACATTCAAAGCAGAGTCTCTCGATTAAGATGGTCTTCTGAAGCCT 2140
Db
481 LeuLysAspArgGluThrPheLysAlaGluSerProAspLysAspGlyLeuLeuLysPro 500
Qy
2141 ACCTGTGGAAGGAAAGTTTCTTCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACA 2200
Db
501 ThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThr 520
Qy
2201 CTCAAAACAGAGTCTCTGATAATGATGCTTCTGAAGCCTACCTGTGGAAGAAAGTT 2260
Db
521 LeuLysAlaGluSerProAspAsnAspGlyLeuLeuLysProThrCysGlyArgLysVal 540
Qy
2261 TCTTCTTCCAAATAAAGCTTTAGAAATTCAGGACAGAGAAACATTCAAAGCAGTCCAGATG 2320
Db
541 SerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaAlaGlnMet 560
Qy
2321 TTCCCATCAGAATCCAAACAAAGGATGATCAAGAAATTTCTTGGGATTTTGAAGTTTC 2380
Db
561 PheProSerGluSerLysGlnLysAspGluGluAsnSerTrpAspPheGluSerPhe 580
Qy
2381 CTTGAGACTCTCTTACAGAAATGATGTGTTTACCCAGGCTACACATCAAAAAGATTC 2440
Db
581 LeuGluThrLeuLeuGlnAsnAspValCysLeuProLysAlaThrHisGlnLysGluPhe 600
Qy
2441 GATACCTTAAAGTGAATAATAGAGAGTCTCCTGATAAAGATGCTTCTTGAAGCCTTACC 2500
Db
601 AspThrLeuSerGlyLysLeuGluGluSerProAspLysAspGlyLeuLeuLysProThr 620
Qy
2501 TGTGGAATGAAAATTTCTTCTTCCAAATAAAGCCTTAGAATGAAGGACAGAGAAATTC 2560
Db
621 CysGlyMetLysIleSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPhe 640
Qy
2561 AAACAGAGGATGTGAGTCTGTAGACTCCACATTCAGTCTTTTGGCAACCGACTACT 2620
Db
641 LysAlaGluAspValSerSerValGluSerThrPheSerLeuPheGlyLysProThrThr 660
Qy
2621 GAAAATTCACAGTCTCAAAAAGTTGAGAAAGCTTTAATCTTACTACCAAGAGGGAGCA 2680
Db
661 GluAsnSerGlnSerThrLysValGluGluAspPheAsnLeuThrLysGluGlyAla 680
Qy
2681 ACAAGACAGTAATCTGCAACACAGGAACTGATATTCGCATTATTGAACGAGTCCACAA 2740
Db
681 ThrLysThrValThrGlyGlnGlnArgAspIleGlyIleIleGluArgGalaProGln 700
Qy
2741 GATCAAAATAAAGATGCCACATCAGAATTAGAAAGAAAAAGAAATACAAAATCAACT 2800
Db
701 AspGlnThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThr 720
Qy
2801 TCAGATTCTGAGATTATCTCTGTGAGTGATACAGAAATTTATGAGTGTTCACCTGAGCT 2860
Db
721 SerAspSerGluIleIleSerValSerAspThrGlnAsnTyrGluCysLeuProGluAla 740
Qy
2861 ACATATCAAAAGAAATAAAGACAAATGGCAAAATAGAGAGTCTCTTCAAAAAGCCT 2920

Db 741 ThrTyrGlnLysGluIleLysThrThrAsnGlyLysIleGluLysProGluLysPro 760
QY 2921 TCTCACTTGAGCTGCCACTGAATGCAAACTCTGTTCCAAATAAAGCGTTAGATGG 2980
Db 761 SerHisPheGluProAlaThrGluMetGlnAsnSerValProAsnLysGlyLeuGluTrp 780
QY 2981 AGAAATAAACAACATTGAGAGCAGATTCAACTPACCCCTATCAAAATCTTGATGACATT 3040
Db 781 LysAsnLysGlnThrLeuArgAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeu 800
QY 3041 CTTCTTGTGAAGAGGAGGAACTTAAAGATGAATGTCGAACAAATTAACAGCAAA 3100
Db 801 ProSerCysGluArgGlyArgGluLeuLysLysAspAsnCysGluGlnIleThrAlaLys 820
QY 3101 ATGGACAAATGAAAAATAGTTTCTGTACTACAAAGAACTGTCAAGGAGGAAAGAA 3160
Db 821 MetGluGlnMetLysAsnLysPheCysValLeuGlnLysGluLeuSerGluAlaLysGlu 840
QY 3161 ATAAATCAAGTTAGAGAACCAAAAGCTAAATGGGAAACAAGAGCTCTGCAGTGTGAGA 3220
Db 841 IleLysSerGlnLeuGluAsnGlnLysAlaLysTrpGluGlnGluLeuCysSerValArg 860
QY 3221 TTGCCTTTAAATCAAGAGAGAGAGAGAGAAATGTCGATATATTAAGAAAAATTT 3280
Db 861 LeuProLeuAsnGlnGluGluGluLysArgArgAsnValAspIleLeuLysGluLysIle 880
QY 3281 AGACCCGAAGAGCAACTTAGGAAAAAGTTAGAAAGTGAACACCAACTTGAACAGACTCTC 3340
Db 881 ArgProGluGlnLeuArgLysLysLeuGluValLysHisGlnLeuGluGlnThrLeu 900
QY 3341 AGAATCAAGATAGAAATGAAGTGAACAGTAAGTAATTTGAATCAGCTTTCTCACACT 3400
Db 901 ArgIleGlnAspIleGluLeuLysSerValThrSerAsnLeuAsnGlnValSerHisThr 920
QY 3401 CATGAAAGTGAATATCTCTTTCATGAAATTTGATGCTGTAAGGAAAGAAATTTCCCATG 3460
Db 921 HisGluSerGluAsnAspPheHisGluAsnLysCysMetLeuLysGluIleAlaMet 940
QY 3461 CTAAACTGGAAGTAGCCACTGAAACATCAACACAGGTGAAGGAAATTAATTAATTT 3520
Db 941 LeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsnLysTyrPhe 960
QY 3521 GAGCAATTAAGTTTTACAGAAAGAGTGTGAATCAATGACCTTAATCAACTGAA 3580
Db 961 GluAspIleLysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLys 980
QY 3581 CAGAAAAAGTAAACAAAAAGGCGATCTCAGTATAGAGACAGCTTAAAGTTCTGACGCA 3640
Db 981 GlnLysThrValThrLysArgAlaSerGlnTyrArgGluGlnLeuLysValLeuThrAla 1000
QY 3641 GAGAACAGATGCTGACTTCTTAATTAAGGAA 3673
Db 1001 GluAsnThrMetLeuThrSerLysLeuLysGlu 1011

RESULT 6

US-10-058-270A-4
; Sequence 4, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Breast Cancer
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472

; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-4
Alignment Scores:
Pred. No.: 2,71e-199 Length: 1341
Score: 3458.00 Matches: 740
Percent Similarity: 65.58% Conservative: 83
Best Local Similarity: 58.96% Mismatches: 159
Query Match: 53.63% Indels: 273
DB: 12 Gaps: 12
US-09-602-362E-26 (1-3673) x US-10-058-270A-4 (1-1341)
QY 349 ATGACAGTAGGGAAGAGCCCGTCAACCTGAACAAAGAGATATGAAGAGAGGACTGCT 408
Db 1 MetThrLysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArgThrAla 20
QY 409 CTACACTGGGCGCTGTCAATGGCCATGCANAAAGTAGTAACATTCTGGTAGACAGAAAG 468
Db 21 LeuHisTrpAlaCysValAsnGlyHisGluGluValValThrPheLeuValAspArgLys 40
QY 469 TGCNCGCTTAATGCTTGAATGGCAAGGAGGACACCTCTGATGAAGGCTCTACAAATGC 528
Db 41 CysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaLeuGlnCys 60
QY 529 GAGAGGAAGCTTTGTGCAAAATTTCTATAGATGCTGTGCTGTGATCTAAATATTATGAGA 588
Db 61 HisGlnGluAla-CysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnLeuValAs 80
QY 589 TGTGTATGGCAACACAGCGCTCTCCATTATGCGTTTATAGTGAAGATTTATTAATGTGGC 648
Db 80 pValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerValValAl 100
QY 649 AACACTCTCTCTATGCTCAGTCTCAGGTGCAAGGTGCAAGGTGCAAGGTGCAAGGTGCA 708
Db 100 aLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuThrProLe 120
QY 709 TTTACTGGCCATACAGAAAGAGCAAGCAAACTGTGGATTTTACTACAAAAATGTC 768
Db 120 uLeuLeuSerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIleLysAsnAl 140
QY 769 AAATGCAAAAGCCATTTAATGAGTCTAAATGCACAGCCCTCATGCTGCCATATGTGAAGG 828
Db 140 aAsnAlaAsnAlaValAsnLysTyrLysCysThrAlaLeuMetLeuAlaValCysHisG 160
QY 829 CTCAATCAGATAGTGGCGATGCTTCTTACGCAAAATGTTGACGCTCTTTGCTGAAGACAT 888
Db 160 ySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAlaAlaAspI 180
QY 889 ACATGGAATAACTGCAGAAAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
Db 180 eCysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIleHisGluG 200
QY 949 ACTTTTGGACATATACGAAAAATTAACCTAAAATCTCAAAATACCAATCCAGAGGAAC 1008
Db 200 nIleMetGluTyrIleArgLysLeuSerLysAsnHisGlnAsnThrAsnProGluGlyTh 220
QY 1009 ATCTACAGGAACACCTGATGAGGCTGCACCTTGGCG----- 1045
Db 220 rSerAlaGlyThrProAspGluAlaAlaProLeuAlaGluArgThrProAspThrAlaG 240

QY 1046 -----GAAAGAACCTGACAC 1062
Db 240 userLeuValGluLysThrProAspGluAlaalaProLeuValGluArgThrProAspTh 260
QY 1063 GCCTCAAGCTTCTGGAAAAACACCTGACGAGCTGCACCTGGTGGAGGAACGTC 1122
Db 260 rAlaGluSerLeuValGluLysThrProAspGluAlaalaSerLeuValGluGlyThrSe 280
QY 1123 TCCCAAAATTCATCTCTGGGAAAGCAACATCTCGAAAGTTTGAAACAGTCAACAGAGA 1182
Db 280 rAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGluGlnSerAlaGluGl 300
QY 1183 AACACTAGGAAATTTTGGAGCCCTACAAAAGAAACATCTGAGAAATTTTCATGSCCA-- 1240
Db 300 uThrProArgGluIleThrSerProAlaLysGluThrSerGluLysPheThrTrpProAl 320
QY 1240 ----- 1240
Db 320 aLysGlyArgProArgLysIleAlaTrpGluLysLysGluAspThrProArgGluIleMe 340
QY 1241 -----GCAAGAAAGATCTAGGAA 1260
Db 340 tSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyA-gProArgL 360
QY 1261 GATCACATGGGAGGAAAAAGAAACATCTGTAAGAGCTGAATCGTGGCAGGAGTAACACC 1320
Db 360 sIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaA-gValThrSe 380
QY 1321 TAATAAACTGAAGTTTGGAAAAAGGAACATCTAATATGATGATGCTCTTACAAAAGA 1380
Db 380 rAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGl 400
QY 1381 AACACTACAAAGCAAGTACAAATCTGGATGTGAGTTCTGTAGAGCCTATATTTCAGTCT 1440
Db 400 uSerSerThrLysAlaSerAlaAsn----- 408
QY 1441 TTTTGGCACACGGCATTTGAAAAATTCACAGTGTACAAAAGTTGAGGAGACTTTAACT 1500
Db 408 ----- 408
QY 1501 TGCTACCAAGATTATCTCTAAGGTGCTGCACAGAATTATACGTCTTTACCTGTAGCTAC 1560
Db 408 ----- 408
QY 1561 ATATCAAAAAGATATCAAAAACATAAATCAAAAATAGAGATCAGATGTTCCCATCAGA 1620
Db 409 -----AspGlnArgPheProSerGl 415
QY 1621 ATCCAAACGAGAGAGATGAGAATATTTCTGGGATTTCTGGAGTCTCTTTGAGAGTTC 1680
Db 415 uSerLysGlnGluAspGluLysrSerCysAspSerArgSerLeuPheGluSerSe 435
QY 1681 TGCAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAAGTAATCGAGATAAATAG 1740
Db 435 rAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnAr 455
QY 1741 AGAAGTAGAAGACTTCTGAGAGCCATCTCCCTCAAGCTGCGGTGAAATCAAAA 1800
Db 455 gGluValGluProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAs 475
QY 1801 GACTGTTCAAATAAAGCTTTGAATTTGAAGATGAACAAACATTCAGAGCAGCTCAGAT 1860
Db 475 nSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspProme 495
QY 1861 GTTCCCATCAGAAATCAAAAAGGACGATGAAGAAAATTTCTGGGATTTCTGAGAGTCC 1920
Db 495 tPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSerLe 515
QY 1921 CTGTGAGAGCTTTTCAGAGAGATGTGTATTTCACCAAGCTACACATCAAAAAGATTT 1980
Db 515 uCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIl 535
QY 1981 CGATACCTTAAGTGGAAAAATTAGAAGAGTCTCTCTGTTAAAGATGGTCTTCTGAGCCTAC 2040

Db 535 eAspLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaTh 555
QY 2041 CTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAAGACAGAGAAACATT 2100
Db 555 rCysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPh 575
QY 2101 CAAGACGAGTCTCTCATAAAGATGGTCTTCTGAAGCCTACCTGTGGAGGAAGATTTC 2160
Db 575 eLysAlaGluProProGlyLysProSerAlaPheGluPro----- 588
QY 2161 TCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGCAGAGTCTCCTGA 2220
Db 589 -----AlaThrGluMetGlnLys----- 594
QY 2221 TAATGATGCTTCTGAGCCTACCTGTGGAAGAAAGTTTCTTCCAAATAAAGCTTT 2280
Db 595 -----SerValProAsnLysAlaLe 601
QY 2281 AGAATTCAAGCAGACAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACA 2340
Db 601 uGluLeuLysAsnGluGlnThrTrpArgAlaAspGluIleLeuProSerGluSerLysGl 621
QY 2341 AAAGGATGATGAAGAAATTTCTGGGATTTTGAGAGTTTCTTGGAGACTCTCTTACAGAA 2400
Db 621 nLysAspTyrGluGluAsnSerTrpAspThrGluSerLeuCysGluThrValSerGlnLy 641
QY 2401 TGATGTGCTTTTACCCCAAGCTACACATCAAAAAGAAATTCGATACATCTTAAGTGAAATTT 2460
Db 641 sAspValCysLeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLe 661
QY 2461 AGAAGACTCTCTCTGATAAAGATGGTCTTCTCAAGCCTACCTGTGGAATGAAATTTCTCT 2520
Db 661 uGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIl 681
QY 2521 TCCAAATAAAGCCTTAGAATTTGAAGCAGAGAAACATTCAAAGCAGAGATGTGAGTTC 2580
Db 681 eProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLy 701
QY 2581 TGTAGAGTCCACATTCAGTCTTTTGGCAAAACCGACTACTGAAAATTTACAGCTTACAAA 2640
Db 701 s---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal-- 715
QY 2641 AGTTGAGGAACACTTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGACTAAGTGCACA 2700
Db 716 -----ProAsnLysAlaLeuGluLeuLy 723
QY 2701 ACAGGAAACGTGATATTGGCATTTATTGAACGAGCTTCCACAAAGATCAAAACAAATAAGATGCC 2760
Db 723 sAsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 735
QY 2761 CACATCAGAATTAGGAAGAAAAGAGATACAAAATCAACTTCAGATTCCTGAGATTATCTC 2820
Db 736 ---SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeuCy 754
QY 2821 TGTGAGTGATACACAGAAATTTAGAGTGTTCACCTGAGGCTACATATCAAAAGAAATATAA 2880
Db 754 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAs 774
QY 2881 GACAAACAAATTCGCAAAATAGAAAGTCT----- 2908
Db 774 pLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCy 794
QY 2908 ----- 2908
Db 794 sArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLy 814
QY 2909 -----CCTCAAAAGCCTTCTCACTTTGAGCCTGCCACTGCAATGCAAAACTCTGT 2958
Db 814 sAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVa 834
QY 2959 TCCAAATAAAGCCTTAGAATGGAAGATAAACAACATTCAGAGCA----- 3004
|||||

Db 240 userLeuValGluLysThrProAspGluAlaAlaProLeuValGluArgThrProAspTh 260
QY 1063 GGCTGAAGCTTCTCTGGAAGAAACACCTGACGAGGTGCACGCTTGGTGAGGAACGTC 1122
Db 260 rAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeuValGluLysThrSe 280
QY 1123 TGCACAAATTCATGCTCTGGGGAAGCAACATCTGGAAGTTTGAACAGTTCACAGACAA 1182
Db 280 rAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGluGlnSerAlaGluG 300
QY 1183 AACACCTAGGAAATTTTGGGCTACAAAGAAACATCTGAGAAATTTTCATGGCCA-- 1240
Db 300 uThrProArgGluIleThrSerProAlaLysGluThrSerGluLysPheThrTrpProAl 320
QY 1240 ----- 1240
Db 320 aLysGlyArgProArgLysIleAlaTrpGluLysLysGluAspThrProArgGluIleMe 340
QY 1241 -----GCAAAAGAAAGACTTAGGAA 1260
Db 340 tSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArgLy 360
QY 1261 GATCACATGGGAGGAAAGAAACATCTGTAAGACTGAATGCGTGGCGAGGTAAACACC 1320
Db 360 sIleAlaTrpGluLysGluThrProValLysThrGlyCysValAlaArgValThrSe 380
QY 1321 TAATAAACTGAAGTTTGGAAAAAGCAACATCTAATATGATTGTCATGCTTACAAAAGA 1380
Db 380 rAsnLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGl 400
QY 1381 AACATCTACAAAAGCAAGTACAAATGTGGATGTGAGTCTCTGTAGACCTATATTCAGTCT 1440
Db 400 userSerThrLysAlaSerAlaAsn----- 408
QY 1441 TTTTGGCACAGGACTATTGAAAATTACAGTGTACAAAAGTTGAGGAAGACTTAATCT 1500
Db 408 ----- 408
QY 1501 TGCTACCAAGATTATCTCTAAGAGTGTGCACAGAATTATACGTGTTTACCTGATGCTAC 1560
Db 408 ----- 408
QY 1561 ATATCAAAAAGATATCAAAACAAATAATCACAAAATAGAAATCAGATGTTCCCATCAGA 1620
Db 409 -----AspGlnArgPheProSerGl 415
QY 1621 ATCCAAACGAGGAGAGTGAAGATATCTTGGGATCTCTGGAGTCTCTTTGAGAGTTC 1680
Db 415 uSerLysGlnGluLysGluLysThrSerCysAspSerArgSerLeuPheGluSerSe 435
QY 1681 TGCAAAAGACTCAAGTCTGTATACCTGAGTCTATGTATCAGAAAGTAATCGAGATAATAG 1740
Db 435 rAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnAr 455
QY 1741 AGAAGTAGAGAGCTTCTCAGAGGACATCTGCCTTCAGGCTCGCGTNGAATGCAAAA 1800
Db 455 gGluValGluLysProLysPheLysProSerAlaPheLysProAlaIleGluMetGlnAs 475
QY 1801 GACTGTTCACAAATPAAAGCCCTTTGAATTGAAGATGAACAAACATTCAGAGCAGCTCAGAT 1860
Db 475 nSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspProMe 495
QY 1861 GTTCCCATCAGATCCAAACAAAGACGATGAGAAAATTTCTGGGATTTCTGAGATGCC 1920
Db 495 tPheProProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSerLe 515
QY 1921 CTGTGAGAGCGGTTTCAGAGAGGATGTGATTATTACCCAAAGCTACACATCAAAAAGAAAT 1980
Db 515 uCysGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluI 535
QY 1981 CGATACCTTAAGTGGAAATTAGAAGAGTCTCTGTTAAAGATGCTTCTCGAAGCCTAC 2040
||| : : : : : |||

Db 535 eAspLysIleAsnGlyLysLeuGluLysSerProAsnLysAspGlyLeuLeuLysAlaTh 555
QY 2041 CTGTGGAAGGAAGTTTCTCTTCCAAATAAAGCCTTAGAATTAAAGGCACAGAAACATT 2100
Db 555 rCysGlyMetLysValSerIleProThrLysAlaLeuGluLysAspMetGlnThrPh 575
QY 2101 CAAGCACAGAGTCTCTGTATAAAGATGCTTCTGAAGCCTACCTCTGGAAGGAAGTTTC 2160
Db 575 eLysAlaGluProProGlyLysProSerAlaPheGluPro----- 588
QY 2161 TCTTCCAAATAAAGCCTTAGAATTAAAGGACAGAGAAACACTCAAGACAGAGTCTCTCTGA 2220
Db 589 -----AlaThrGluMetGlnLys----- 594
QY 2221 TAATCATGCTCTCTGAAGCCTTCTGTGAAGGAAAGTTTCTCTCCAAATAAAGCTTT 2280
Db 595 -----SerValProAsnLysAlaLe 601
QY 2281 AGAATTGAAGCACAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACA 2340
Db 601 uGluLeuLysAsnGluGlnThrTrpArgAlaAspGluIleLeuProSerGluSerLysGl 621
QY 2341 AAAGATGATGAAGAAATTTCTGGGATTTTGAGATTTTCCCTTGAGACTCTCTTACAGAA 2400
Db 621 nLysAspTyrGluGluAsnSerTrpAspThrGluSerLeuCysGluThrValSerGlnLy 641
QY 2401 TGATGTGTGTTTACCAGGCTACACATCAAAAAGAAATTCGATACCTTAAGTGAATAAT 2460
Db 641 sAspValCysLeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLe 661
QY 2461 AGAAGAGTCTCTGTATAAAGATGCTTCTGAAGCCTACCTGTGGAATGAAAATTTCTCT 2520
Db 661 uGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIl 681
QY 2521 TCCAAATAAAGCCTTAGAATTGAAGGACAGAGAACATTCAAAGCAGAGGATGTGAGTTC 2580
Db 681 eProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLy 701
QY 2581 TGTAGATGCCAATTCAGTCTTTTGGCAACCGAGTACTCAAAATTCACAGTCTACAAA 2640
Db 701 s---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal-- 715
QY 2641 AGTTGAGGAGACTTTAATCTTACTACCAAGGAGGAGCAACAAGACAGTAACTGGACA 2700
Db 716 -----ProAsnLysAlaLeuGluLeuLy 723
QY 2701 ACAGGAACGTGATATTGGCATTATTGAACGAGCTCCACAAGATCAACAATAAGATGCC 2760
Db 723 sAsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 735
QY 2761 CACATCAGAATTAGGAAGAAAGAGATACAAATCAACTTCAGATTCAGATTCTGAGATTCTC 2820
Db 736 ---SerGluSerLysGlnLysAspTyrGluSerSerTrpAspSerGluSerLeuCy 754
QY 2821 TGTGAGTGATACACAGAAATTTATGATGTTTACCTGAGGCTACATATCAAAAAGAAATAAA 2880
Db 754 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAs 774
QY 2881 GACAACAATAGCAAAATAGAAAGATCT----- 2908
Db 774 pLysIleAsnGlyLysLeuGluLysSerProAspAsnAspGlyPheLeuLysAlaProCy 794
QY 2908 ----- 2908
Db 794 sArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLy 814
QY 2909 -----CCTGAAAGCCTTCTCATTGTAGCCTGCCACTGAAATGCAAAACTCTGT 2958
Db 814 sAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVa 834
QY 2959 TCCAAATAAAGCCTTAGAATGGAAGAAATAACAACATTGAGAGCCA----- 3004
Db 834 lProAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePr 854
||| : : : : : |||

```

QY 3004 ----- 3004
Db oSerGluSerLysGlnLysValGluGluAenSerTrpAspSerGluSerLeuArgGl 874
QY 3004 ----- 3004
Db uThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys 894
QY 3005 -----CAATCAACTACCCCTATCAAAAATCTTGATGACCTTCCTTC 3045
Db sIleSerGlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHis 914
QY 3046 TTGTGAAGAGGAGGAACTTAAAGAAAGTAACTGTGAACAAATACACGCAAAATGGA 3105
Db rCysGluArgAlaArgGluLeuGlnLysAspHisCysGluGlnArgThrGlyLysMetGl 934
QY 3106 ACAAATGAAAAATAAGTTTGTCTACTACAAAGAACTGTGAGAAGCGAAGAAATAAA 3165
Db uGlnMetLysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluLeu 954
QY 3166 ATCACAGTTAGAGAACCAAAAGCTAAATGCGAACAGAGCTCTGCAGTGTGAGATTGCC 3225
Db sSerGlnLeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerValArgLeuTh 974
QY 3226 TTTAAATCAAGAAGAGAGAGAGAAATGTCATATATTAAAGAAAAAATAGACC 3285
Db rLeuAsnGlnGluGluLysArgAsnAlaAspIleLeuAsnGluLysIleArgGl 994
QY 3286 C-----CAAGAGCACTTAGGAAAAGTTAGAGTGAACACCACTTCA 3330
Db uGluLeuGlyArgIleGluGlnHisArgLysGluLeuGluValLysGlnGlnLeuGl 1014
QY 3331 ACAGACTCTCAGATAACAAGATAGAAATGAAAAGTGTAAACAGTAATTTGAATCAGGT 3390
Db uGlnAlaLeuArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGlnVa 1034
QY 3391 TTCTCACACTCATGAAGTGAATAATGATCTCTTTTCATGAAAATGATGATGTGAAAAGGA 3450
Db lSerHisThrHisGluAsnGluAsnTrpLeuLeuHisGluAsnCysMetLeuLysLysGl 1054
QY 3451 AATTGCCATCTAAACCTGAACTAGCCACACTGAAACATCAACACAGCTGAAGGAAA 3510
Db uIleAlaMetLeuLysLeuGluLeuAlaThrLeuLysHisGlnTrpGlnGluLysGluAs 1074
QY 3511 TAAATCTTTGAGACATTAAATTTTACAAGAAAGATGCTGAACTTCAATGACCTT 3570
Db nLysTyPheGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLe 1094
QY 3571 AAAATCAAAACAGAAACAGTAAACAAAGGGATCTCAGTATAGAGAGCAGCTTAAAGT 3630
Db uLysLeuLysGluGluSerLeuThrLysArgAlaSerGlnTrpSerGlyGlnLeuLysVa 1114
QY 3631 TCTGACGGCAGAGAACGATGCTGACTTCTAAATTCAGGAA 3673
Db lLeuIleAlaGluAsnThrMetLeuThrSerLysLeuLysGlu 1128

```

RESULT 8

```

US-10-076-622-565
; Sequence 565, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076.622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ. ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565

```

```

; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-076-622-565

```

```

Alignment Scores:
Pred. No.: 2,71e-199 Length: 1341
Score: 3458.00 Matches: 740
Percent Similarity: 65.58% Conservative: 83
Best Local Similarity: 58.96% Mismatches: 159
Query Match: 53.63% Indels: 273
DB: 14 Gaps: 12

```

US-09-602-362E-26 (1-3673) x US-10-076-622-565 (1-1341)

```

QY 349 ATGACAGTAGGGAAGAGCCCGTCAACCTGAACAAAGAGATATGAAGAAGAGGACTGCT 408
Db 1 MetThrLysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArgThrAla 20
QY 409 CTACACTGGGCGCTGTGTCATATGCGCATGCGANAGTAGTAACTTCTGGTAGACAGAAG 468
Db 21 LeuHisTrpAlaCysValAsnGlyHisGluValValThrPheLeuValAspArgLys 40
QY 469 TGCNCGCTTAATGTCCTTGTATGGGGAAGGAGGACACCTCTGATGAAGGCTTACAAATGC 528
Db 41 CysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaLeuGlnCys 60
QY 529 GAGAGGGAAGCTTTGTGCAATATTCATAGATGCTGCTGCTGATCTTAATATTATGTAGA 588
Db 61 HisGlnGluAla-CysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnLeuValAs 80
QY 589 TGTGTATGGCAACACGCGCTCTCCATTATGCGTTTATAGTGAGAATTTATTATGTGGGC 648
Db 80 pValTrpGlyAsnMetAlaLeuHisTrpAlaValTrpSerGluLeuLeuSerValValAl 100
QY 649 AACACTGTGTCCTATGTCAGTCATCGAGGTGCAAAACAGGCTAGCCTCAACCCCT 708
Db 100 aLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuThrProLe 120
QY 709 TTTACTGCGCATACAGAAAGCAACCAAGCAACTGTGGAATTTTACTACAAAAATGC 768
Db 120 uLeuLeuSerIleThrLysArgSerGluGlnIleValGluPheLeuLeuLysAsnAl 140
QY 769 AAATGCAAAACCAATTAATGAGTCTAAATGACACAGCCCTCATGCTGCCATATGTGAAGG 828
Db 140 aAsnAlaAsnAlaValAsnLysTyLysCysThrAlaLeuMetLeuAlaValCysHisGl 160
QY 829 CTCATCAGAGATAGTCGCGCATGCTTTCAGCAAAATGTTGACGCTTTGCTGAGACAT 888
Db 160 ySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAlaAlaAspIl 180
QY 889 ACATGGAATACTCAGAACCTTATGCTGCTGCTCGTGGAGTTAATTAATTCATTTCAACA 948
Db 180 eCysGlyValThrAlaGluHisTyThrAlaValThrCysGlyPheHisIleHisGluGl 200
QY 949 ACTTTGGAACATATACGAAAATTTACCTAAATCTCAAAATACCAATACCAAGGAAC 1008
Db 200 nIleMetGluTyTrileArgLysLeuSerLysAsnHisGlnAsnThrAsnProGluGlyTh 220
QY 1009 ATCTACAGAACACCTGATGAGGCTGCACCCCTCGCG----- 1045
Db 220 rSerAlaGlyThrProAspGluAlaAlaProLeuAlaGluArgThrProAspThrAlaGl 240
QY 1046 -----GAAAGAACCTGCAC 1062
Db 240 uSerLeuValGluLysThrProAspGluAlaAlaProLeuValGluArgThrProAspTh 260
QY 1063 GCCTGAAAGCTTGTGGNAAAAACCTGACGAGCTGCACGCTTGTGTGAGGGAAGCTC 1122
Db 260 rAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeuValGluGlyThrSe 280
QY 1123 TGCCAAAATTCAAATGCTTGGGGAAGCAACATCTGGAAAGTTTGAACAGCTCAACAGAAGA 1182

```

Db 280 rAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGluGlnSerAlaGluGl 300
QY 1183 AACACCTAGGAAATTTTGGAGCTACAAAGAAACATCTGAGAAATTTTCATGGCCA-- 1240
Db 300 uThrProArgGluIleThrSerProAlaLysGluThrSerGluLysPheThrTrpProAl 320
QY 1240 ----- 1240
Db 320 aLysGlyArgProArgLysIleAlaTrpGluLysLysGluAspThrProArgGluIleMe 340
QY 1241 -----GCAAAAGAAAGATCTAGGAA 1260
Db 340 tSerProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArgLys 360
QY 1261 GATCACAATGGAGGAGAAAGAAACATCTGTAAGACTGAATGCGTGGCAGGAGTAACACC 1320
Db 360 sIleAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThrSe 380
QY 1321 TAATAAAGCTGAAGTTTGGAAAGAGAACATCTAATATGATTGCGATGCTCTACAAAAGA 1380
Db 380 rAenLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGl 400
QY 1381 AACATCTACAAAAGCAAGTACAAATGTGGATGTGAGTTCTGTAGAGCCTATATTCAGTCT 1440
Db 400 uSerSerThrLysAlaSerAlaAsn----- 408
QY 1441 TTTTGGCACACGACTATTGAAAATTACAGTGTACAAAAGTTGAGGAAGACTTTAATCT 1500
Db 408 ----- 408
QY 1501 TGTACCAAGATTATCTTAAGAGTGTGCACAGAATTATACGTGTTTACCTGATGCTAC 1560
Db 408 ----- 408
QY 1561 ATATCAAAAAGATATCAAAACAATAATCAAAAATAGAAGATCAGATGTTCCCATCAGA 1620
Db 409 -----AspGlnArgPheProSerGl 415
QY 1621 ATCCAAACGAGAGGAGATGAAGAATATCTCTGGGATCTGGAGTCTCTTTGAGAGTTC 1680
Db 415 uSerLysGlnGluLysGluLysGluLysSerCysAspSerArgSerLeuPheGluSerSe 435
QY 1681 TGCAAGACTCAAGTGTGTATACCTGAGTCTATGTATATCAGAAAGTAAATCGAGATAATAG 1740
Db 435 rAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysValMetGluIleAsnAr 455
QY 1741 AGAAGTAGAAGACTCTTCAGAGGACATCTGCTTCAAGCTGCGGTGAAATGCARAA 1800
Db 455 gGluValGluLysProProLysLysProSerAlaPheLysProAlaIleGluMetGlnAs 475
QY 1801 GACTGTTCCAAATAAAGCCTTTCAATTGAAGATGAACAAACATTCAGAGCAGCTCAGAT 1860
Db 475 nSerValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaaspProMe 495
QY 1861 GTTCCCATCAGAATCCAAACAAAGGACGATGAGAAATTTCTGGGATCTTGAGAGTCC 1920
Db 495 tPheProGluSerLysGlnLysAspTyrGluGluAsnSerTrpAspSerGluSerLe 515
QY 1921 CTGTGAGAGCGTTTTCAGAGAGATGTTATTTACCACAAAGCTACACATCAAAAAGAAAT 1980
Db 515 uCysGluThrValSerGlnLysaspValCysLeuProLysAlaThrHisGlnLysGluIle 535
QY 1981 CGATACCTTAAGTGGAAATATAGAGAGTCTCTGTTAAAGATGTCCTCTGAGCCTAC 2040
Db 535 eAspLysIleAsnGlyLysLeuGluSerProAsnLysAspGlyLeuLeuLysAlaTh 555
QY 2041 CTGTGGAGGAAGTTCTCTTCCAAATAAAGCCTTAGAATTAAGGACGAGAGAAACATT 2100
Db 555 rCysGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPh 575
QY 2101 CAAAGCAGAGTCTCTGATAAAGATGCTCTTCTGAAAGCCTACCTGTGGAAGGAAAGTTTC 2160
Db 575 eLysAlaGluProProGlyLysProSerAlaPheGluPro----- 588

QY 2161 TCTTCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACACTCAAGACAGAGTCTCCTGA 2220
Db 589 -----AlaThrGluMetGlnLys----- 594
QY 2221 TAATGATGGTCTTCTGAAGCCTACCTGTGGAAGGAAAGTTTCTTCCAAATAAAGCTTT 2280
Db 595 -----SerValProAsnLysAlaLe 601
QY 2281 AGAATTGAAGCAGAGAGAAACATTCAAAGCAGCTCAGATGTTCCCATCAGAAATCCAAACA 2340
Db 601 uGluLeuLysAsnGluGlnThrTrpArgAlaAspGluIleLeuProSerGluSerLysGl 621
QY 2341 AAAGGATGATCAAGAAATCTCTGGGATTTTGAGAGTTTCTTTCAGACTCTCTTACAGAA 2400
Db 621 nLysAspTyrGluGluAsnSerTrpAspThrGluSerLeuCysGluThrValSerGlnLys 641
QY 2401 TGATGTGTGTTTACCACAGGCTACACATCAAAAAGAATTCGATACCTTAAGTGGAAAT 2460
Db 641 sAspValCysLeuProLysAlaAlaHisGlnLysGluIleaspLysIleAsnGlyLysLe 661
QY 2461 AGAAGAGTCTCTGATAAAGATGCTCTTCTGAAAGCCTACCTGTGGAATGAAAAATTTCTCT 2520
Db 661 uGluGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIl 681
QY 2521 TCCAAATAAAGCCTTAGAATTAAGGACAGAGAAACATTCAAAGCAGAGGATGTGAGTTC 2580
Db 681 eProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLys 701
QY 2581 TGTAGAGTCCACATTCAGTCTTTTGGCAACCGACTACTGAAAATTCACAGTCTCAAA 2640
Db 701 s---ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal-- 715
QY 2641 AGTTGAGGAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACCTGGACA 2700
Db 716 -----ProAsnLysAlaLeuGluLeuLys 723
QY 2701 ACAGGACGTGATATTGGCATTATTGAACGAGCTCCACAGATCAACAAATAAGATGCC 2760
Db 723 sAsnGluGlnThrLeuArgAlaAspGluIleLeuPro----- 735
QY 2761 CACATCAGAAATAGGACAGAAAGAGATACAAAATCAACTTCAGATTTCTGAGATTATCTC 2820
Db 736 ---SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLeuCy 754
QY 2821 TGTGAGTGATACACAGAAATTTATGAGTGTTTACCTGAGGCTACATATCAAAAAGAAATAA 2880
Db 754 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAs 774
QY 2881 GACAAACAATGGCAAAATAGAAAGATCT- 2908
Db 774 pLysIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCy 794
QY 2908 ----- 2908
Db 794 sArgMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLys 814
QY 2909 -----CCTGAAAAGCCTTCTCATTGAGCCTGCCACTGAAATGCAAAACTCTGT 2958
Db 814 sAlaGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerVa 834
QY 2959 TCCAAATAAAGCCTTAGAATGGAAGATTAACAAACATTTGAGAGCA----- 3004
Db 834 lProAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPhePr 854
QY 3004 ----- 3004
Db 854 oSerGluSerLysGlnLysLysValGluGluAsnSerTrpAspSerGluSerLeuArgGl 874
QY 3004 ----- 3004
Db 874 uThrValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLys 894

Db 249 rLeuValGluLysThrProAspGluAlaalaProLeuValGluArgThrProAspThrAl 269
QY 1066 TGAAGCTTGGCTGGAAAAACACCTGACGAGGCTGCACGCTTGGTGGAGGAACGCTGTC 1125
Db 269 aGluSerLeuValGluLysThrProAspGluAlaalaSerLeuValGluGlyThrSerAs 289
QY 1126 CAAATTCATATGCTGGGAAGCAACATCTGGAAGTTTGAACAGTCAACAGAGAAGAAC 1185
Db 289 pLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGluGlnSerAlaGluGluTh 309
QY 1186 ACCTAGGAATAATTTGAGCGCTACAAAAGAACATCTGAGAATTTTCATGGCCA- 1240
Db 309 rProArgGluIleThrSerProAlaLysGluThrSerGluLysPheThrTrpProAlaLys 329
QY 1240 - 1240
Db 329 sGlyArgProArgLysIleAlaTrpGluLysLysGluAspThrProArgGluIleMetSe 349
QY 1241 - 1241
Db 349 rProAlaLysGluThrSerGluLysPheThrTrpAlaLysGlyArgProArgLysIle 369
QY 1264 CACATGGGAGGAAAAAGAACATCTGTAAGACTGGAATCGCTGGCAGGAGTAACACCTAA 1323
Db 369 eAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThrSerAs 389
QY 1324 TAAACTGAAGTTTGGAAAAAGAACATCTAATATGATGCTGCTACAAAAGAAC 1383
Db 389 nLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGluSe 409
QY 1384 ATCTACAAAGCAAGTACAAATGTGGATGTGAGTCTGTAGAGCCTATATTTCAGTCTTTT 1443
Db 409 rSerThrLysAlaSerAlaAsn- 416
QY 1444 TGSCACACGAGCTATTGAAAAATTCACAGTGTACAAAGTTGAGGAAGCTTTAACTTTCG 1503
Db 416 - 416
QY 1504 TACCAAGATTATCTTAAGAGTCTGCACAGAATTATACGTGTACCTGTATGCTACATA 1563
Db 416 - 416
QY 1564 TCAAAAGATATCAAAACAAATAATCAAAATAGAGATCAGATGTTCCCATCAGATC 1623
Db 417 - 417
QY 1624 CAAACGAGAGAGATGAAGAAATATCTTGGGATCTGGGAGTCTCTTTGAGAGTTCTGC 1683
Db 424 rLysGlnGluGluAspGluLysThrSerCysAspSerArgSerLeuPheGluSerSerAl 444
QY 1684 AAAGACTCAAGTGTGTATACCTGAGTCTATGTATCAGAAAGTAATGGAGATAAATAGAGA 1743
Db 444 aLysIleGlnValCysIleProGluSerIleThrGlnLysValMetGluIleAsnArgL 464
QY 1744 AGTAGAAGAGCTTCCTGAGAGCCATCTGCTTCAAGCTCGCGTNGAAATCAAAAGAAC 1803
Db 464 uValGluGluProProLysPheSerProAlaPheLysProAlaIleGluMetGlnAsnSe 484
QY 1804 TGTTCCAAATAAGCTTTGAATTTGAAGAAATGAACAAACATTGAGAGCAGCTCAGATGTT 1863
Db 484 rValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeuArgAlaAspProMetPhe 504
QY 1864 CCATCAGATCAAAACAAAGGACGATGAAGAAATTTCTGGGATTTCTGAGAGTCCCTG 1923
Db 504 eProProGluSerLysGlnLysAspThrGluGluAsnSerTrpAspSerGluSerLeuCy 524
QY 1924 TGAGAGGTTTCACAGAGGATGTGTATTACCAAGACTACATCAACAAAGAAATTCGA 1983
Db 524 sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAs 544
QY 1984 TACCTTAAGTGGAAAAATTAAGAGAGTCTCTGTTAAAGATGGTCTTCTGAGCCTACCTG 2043
Db 544 pLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrCy 564

QY 2044 TGAAGGAAAGCTTTCTCTTCCAAATAAAGCCTTGAATTTAAAGACAGAGAAACATTCAA 2103
Db 564 sGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPheLy 584
QY 2104 AGCAGAGTCTCTCTGATTAAGATGTCTTCTGAAGCCTACCTGGAAGAAAGTTTCTCT 2163
Db 584 sAlaGluProGlyLysProSerAlaPheGluPro- 596
QY 2164 TCCAAATAAAGCCCTTAGAATTTAAAGGACAGAGAAACACTCAAAGCAGAGTCTCTCTGATAA 2223
Db 597 - 602
QY 2224 TGATGGTCTTCTGAAGCCTACCTGTGAAGAAAGTTTCTTCCAAATAAAGCTTTAGA 2283
Db 603 - 610
QY 2284 ATTGAAGCAGACAGAGAAACATTCAAAGCAGCTCAGATGTCCCATCACAATCCAAACAAA 2343
Db 610 uLeuLysAsnGluGlnThrTrpArgAlaAspGluIleLeuProSerGluSerLysGlnLy 630
QY 2344 GGATGATGAAGAAATTTCTTGGGATTTTGAGAGTTTCTCTGAGACTCTCTTACAGAAATGA 2403
Db 630 sAspTy-GluGluAsnSerTrpAspThrGluSerLeuCysGluThrValSerGlnLysAs 650
QY 2404 TGTCTGTTTACCCCAAGCTACACATCAAAAAGATTCGATACCTTAAGTGGAAATTAGA 2463
Db 650 pValCysLeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGl 670
QY 2464 AGAGTCTCTCTGATAAAGATGCTTCTTGAAGCCTACCTGTGGAAATGAAATTTCTCTCC 2523
Db 670 uGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIlePr 690
QY 2524 AAATAAGCCTTAGAATTTGAAGCAGAGAAACATTCAAAGCAGAGGATGTGATTTCTGT 2583
Db 690 oThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLys- 709
QY 2584 AGAGTCCACATTCAGTCTTTTGGCAACCCACTACTGAAAATTCACAGCTCAACAAAGT 2643
Db 710 -ProSerAlaPhe- 723
QY 2644 TGAGGAAGACTTTAATCTTACTACCAAGGAGGAGCAACAAAGACAGTAACTGAGCAACA 2703
Db 724 - 732
QY 2704 GGAAGCTGATATTGGCATTATTGAACGAGCTCCACAGATCAACAAATAAGATGCCAC 2763
Db 732 nGluGlnThrLeuArgAlaAspGluIleLeuPro- 743
QY 2764 ATCAGAATTAGGAAGAAAGAAAGATACAAATCAAAATTCAGATTCTGAGATTATCTCTGT 2823
Db 744 -SerGluSerLysGlnLysAspTyThrGluSerSerTrpAspSerGluSerLysGly 763
QY 2824 GAGTCATACAGAAATTTAGTGTCTTACCTGAGGCTTACATATCAAAAGAAATAAGAC 2883
Db 763 uThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAspLy 783
QY 2884 AACAAATGGCAAAATAGCAAGTCT- 2908
Db 783 sIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCysAr 803
QY 2908 - 2908
Db 803 gMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAl 823
QY 2909 - 2961
Db 823 aGluProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerValPr 843
QY 2962 AAATAAAGGCTTAGAATGGAGAAATAAACAACATTGAGAGCA- 3004
Db 843 oAsnLysAlaLeuGluLysAsnGluGlnThrLeuArgAlaAspGlnMetPheProSe 863

Qy	1186	ACCTAGGAAATTTTGAGGCCTACAAAGAAACATCTCGAGAAATTTTTCATGGCCA	-----	1240
Db	309	rProArgGluIleThrSerProAlaLysGluThrSerGluLysPheThrTrpProAlaLys	:::	
Qy	1240	-----	-----	1240
Db	329	sGlyArgProArgLysIleAlaTrpGluLysLysGluAspThrProArgGluIleMetSe		349
Qy	1241	-----	-----GCAAAAGAAAGATCTAGAAAGAT	1263
Db	349	rProAlaLysGluThrSerGluLysPheThrTrpAlaAlaLysGlyArgProArgLysIle		369
Qy	1264	CACATGGGAGGAAAAAGAAACATCTGTAAAGACTCAATCGTCGACAGAGTAACACCTAA		1323
Db	369	eAlaTrpGluLysLysGluThrProValLysThrGlyCysValAlaArgValThrSerAs	:::	389
Qy	1324	TAAACCTCAAGTTTTGGAAAAAGGAACATCTTAATATGATTCGATCTCTCTACAAAGAAAC		1383
Db	389	nLysThrLysValLeuGluLysGlyArgSerLysMetIleAlaCysProThrLysGluSe		409
Qy	1384	ATCTACAAAGCAAGTCAAAATCGGATCGATCTCTGTAGAGCTATATTCAGTCTCTTT		1443
Db	409	rSerThrLysAlaSerAlaAsn		416
Qy	1444	TGGCACACGGACTATTGAAATTCACAGTGTACAAAGATTGAGGAAGACTTTAATCTTGC	-----	1503
Db	416	-----	-----	416
Qy	1504	TACCAAGATATCTTAAGAGTGTGCACAGAAATTACGTGTTTACCTGATGCTACATA	-----	1563
Db	416	-----	-----	416
Qy	1564	TCAAAAGATATCAAAACAATAAATCAAAATAGAAGATCAGATGTTCCCATCAGAATC		1623
Db	417	-----	-----AspGlnArgPheProSerGluSe	424
Qy	1624	CAACGAGAGGAAGATGAAGAATATCTTGGGATCTCTGGAGTCTCTTTGAGTCTTCTGC		1683
Db	424	rLysGlnGluLysAspGluGluTySerCysAspSerArgSerLeuPheGluSerSerAl		444
Qy	1684	AAAGACTCAAGTGTGTATACCTGAGTCTATCTATCAGAAAGTAATAGAGATAAATAGAGA		1743
Db	444	AlaLysIleGlnValCysIleProGluSerIleTyGlnLysValMetGluIleAsnArgL		464
Qy	1744	AGTAGAGAGAGCTTCTCGAGAAGCCATCTGCCCTTCAAGCCTCGCCCTGNAATGCAAAAGAC		1803
Db	464	uValGlnGluProLysLysProSerAlaPheLysProAlaIleGluMetGlnAsnSe		484
Qy	1804	TGTTCCCAATTAAGCCTTTGAATTGAAGATGAAGAAATGAACAACATGTAGAGCAGCTCAGATGTT		1863
Db	484	rValProAsnLysAlaPheGluLeuLysAsnGlnThrLeuArgAlaAspProMetPh		504
Qy	1864	CCCATCAGAAATCCAAACAAAAGGACGATGAAGAAATTTCTGGGATCTCTGAGATCCCTGT		1923
Db	504	eProProGluSerLysGlnLysAspTyGluGluAsnSerTrpAspSerGluSerLeuCy		524
Qy	1924	TGACAGCGTTTCACAGAAGGATGTATTATTACCAAGCTACACATCAAAAAGAAATTCGA		1983
Db	524	sGluThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAs		544
Qy	1984	TACCTTAAGTGAAATTAGAAGAGTCTCTCTGTTTAAGATGGTCTTCTGGAAGCCCTACCTGT		2043
Db	544	pLysIleAsnGlyLysLeuGluGluSerProAsnLysAspGlyLeuLeuLysAlaThrCy		564
Qy	2044	TGGAAGGAAAGTTTCTCTTCCAAATTAAGCCTTAGAATTAAGACACAGAGAAACATTCAA		2103
Db	564	sGlyMetLysValSerIleProThrLysAlaLeuGluLeuLysAspMetGlnThrPheLys		584
Qy	2104	AGCAGAGTCTCTGTATAAGATGGTCTTCTGGAAGCCTACCTGTGTGGAAGAAAGTTTCTCT		2163
Db	584	sAlaGluProGluLysProSerAlaPheGluPro		596

2164	QY	TCCAAATAAAGCCCTTAGAATTAAGGACGAGAAAACACTCAAGACGAGTCTCTCTGATAA	2222
597	Db	-----AlaThrGluMetGlnLys-----	602
2224	QY	TGATGGTCTTCTGAAGCCTACCTGTGGAAGGAAAGTTTCTCTTCCAAATAAAGCTTTAGA	2283
603	Db	-----SerValProAsnLysAlaLeuGI	610
2284	QY	ATTGAAGGACGAGAGAAAACATTCAAAGCAGCTCAGATGTTTCCCATCAGAAATCCAAACAAA	2343
610	Db	uLeuLysAsnGluGlnThrTrpArgAlaAspGluLeuProSerGluSerLysGlnLys	630
2344	QY	GGATGATGAAGAAAATCTTGGGATTTTGAGAGTTTCCCTTGAGACTCTCTTACAGAATGA	2403
630	Db	sAspTrpGluGluAsnSerTrpAspThrGluSerLeuCysGluThrValSerGlnLysAs	650
2404	QY	TGTGTGTTTACCCAAGCCTACACATCAAAAGAAATTCGATACCTTAAAGTCGAAAATTAGA	2463
650	Db	pValCysLeuProLysAlaAlaHisGlnLysGluLeuAspLysIleAsnGlyLysLeuGI	670
2464	QY	AGAGTCTCCTGATAAAGATGGTCTTCTGGAACCTACTCCTGTGGAAATGAAAATTTCTCTTC	2523
670	Db	uGlySerProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIlePr	690
2524	QY	AAATAAAGCCCTTAGAATTTGAAGACACAGAGAAAACATTCAAACGACGAGATGTGAGTTCTGT	2583
690	Db	oThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLys--	709
2584	QY	AGAGTCACAAATTCAGTCTTTTGGCAACCGACTACTGAAAATTCACAGTCTACAAAAGT	2643
710	Db	ProSerAlaPhe-----GluProAlaIleGluMetGlnLysSerVal-----	723
2644	QY	TGAGGAAGACTTTTAATCTTACTACCAAGGAGGAGCAACAAACACACATTAAGTGCACAACA	2703
724	Db	-----ProAsnLysAlaLeuGluLeuLysAs	732
2704	QY	GGAAAGCTGATATTGGCATTATTGAACGAGCTCCACAGATCAACAATAAGATGCCCAC	2763
732	Db	nGluGlnThrLeuArgAlaAspGluIleLeuPro-----	743
2764	QY	ATCAGAATTAGGAAGAAAGATACAAATCAACTTCAGATTCTCAGATTATCTCTGT	2823
744	Db	-SerGluSerLysGlnLysAspTyrGluGluSerSerTrpAspSerGluSerLysGluGI	763
2824	QY	GAGTGATACACAGAAATATGAGTGTTTTACCTGAGGCTACATATCAAAAAGAAATAAAGAC	2883
763	Db	uThrValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAspLys	783
2884	QY	AACAAATGGCAAAATAGAGAGTCT-----	2908
783	Db	sIleAsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCysAr	803
2908	QY	-----	2908
803	Db	gMetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAl	823
2909	QY	-----CCTGAAAAGCCTTCTCACTTTGAGCCTGCCACTGAAATGCAAACTCTGTCTCC	2961
823	Db	aGluProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerValPr	843
2962	QY	AAATAAAGGCTTAGAATGGGAAGATAAACAACAATTGAGAGCA-----	3004
843	Db	oAsnLysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPheProSe	863
3004	QY	-----	3004
863	Db	xGluSerLysGlnLysLysValGluGluAsnSerTrpAspSerGluSerLeuArgGluTh	883
3004	QY	-----	3004
883	Db	rValSerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLysII	903
3005	QY	-----GATTCAACTACCCCTCAAAAATCTTGATGCACCTCTCTCTTG	3048

